IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re U.S. Patent Application of

KASAKO

Application Number: 10/765,128 **Art Unit 2188**

Filed: January 28, 2004

For: DATA PROCESSING SYSTEM HAVING A PLURALITY OF STORAGE SYSTEMS

Atty Docket No. WILL.0001

Honorable Assistant Commissioner for Patents
Washington, D.C. 20231

REQUEST FOR PRIORITY
UNDER 35 U.S.C. § 119
AND THE INTERNATIONAL CONVENTION

Sir:

In the matter of the above-captioned application for a United States patent, notice is hereby given that the Applicant claims the priority date of December 15, 2003, the filing date of the corresponding Japanese Patent Application 2003-416414.

A certified copy of Japanese Patent Application 2003-416414 is being submitted herewith. Acknowledgment of receipt of the certified copy is respectfully requested in due course.

Respectfully submitted,

Stanley P. Fisher
Registration Number 24,344

Juan Carlos A. Marquez
Registration Number 34,072

REED SMITH LLP
3110 Fairview Park Drive
Suite 1400
Falls Church, Virginia 22042
(703) 641-4200
June 18, 2004

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日 Date of Application:	２００３年１２月１５日
出願番号 Application Number:	特願２００３－４１６４１４
[ST. 10/C]:	[ＪＰ２００３－４１６４１４]
出願人 Applicant(s):	株式会社日立製作所

２００４年　１月２７日

特許庁長官
Commissioner,
Japan Patent Office

今井康夫

出証番号　出証特２００４－３００２９６５

【書類名】　特許願
【整理番号】　340301444
【あて先】　特許庁長官殿
【国際特許分類】　G06F 12/16
G06F 03/06
【発明者】
【住所又は居所】　神奈川県小田原市中里３２２番２号　株式会社日立製作所ＲＡＩＤシステム事業部内
【氏名】　加迫　尚久
【特許出願人】
【識別番号】　000005108
【氏名又は名称】　株式会社日立製作所
【代理人】
【識別番号】　100095371
【弁理士】
【氏名又は名称】　上村　輝之
【選任した代理人】
【識別番号】　100089277
【弁理士】
【氏名又は名称】　宮川　長夫
【選任した代理人】
【識別番号】　100104891
【弁理士】
【氏名又は名称】　中村　猛
【手数料の表示】
【予納台帳番号】　043557
【納付金額】　21,000円
【提出物件の目録】
【物件名】　特許請求の範囲　1
【物件名】　明細書　1
【物件名】　図面　1
【物件名】　要約書　1
【包括委任状番号】　0110323

【書類名】特許請求の範囲
【請求項１】
上位装置に通信可能に接続された第一の記憶システムと、
前記第一の記憶システムにそれぞれ通信可能に接続された第二及び第三の記憶システムと備え、
（１）　前記第一の記憶システムは、
前記上位装置から送られたデータが格納される第一のデータ記憶領域と、
前記第一の記憶領域に格納されたデータの複製を生成するために使用されるジャーナルが格納される第一のジャーナル記憶領域と、
前記上位装置から送られたデータを前記第一のデータ記憶領域に書き、前記第一のデータ記憶領域に書かれたデータのジャーナルを前記第一のジャーナル記憶領域に書き、前記第二及び第三の記憶システムのそれぞれからの要求に応じて前記第一のジャーナル記憶領域内の前記ジャーナルを前記第二及び第三の記憶システムにそれぞれ送るようになった第一の制御装置と
を有しており、
（２）　前記第二の記憶システムは、
前記第一のデータ記憶領域内のデータの複製が格納される第二のデータ記憶領域と、
前記ジャーナルが格納される第二のジャーナル記憶領域と、
独自にスケジュールされたジャーナルリードのタイミングで前記第一の記憶システムから前記ジャーナルを読み、読まれたジャーナルを前記第二のジャーナル記憶領域に書き、そして、独自にスケジュールされたリストアのタイミングで前記第二のジャーナル記憶領域内の前記ジャーナルに基づいて前記第一のデータ記憶領域内のデータの複製を生成して前記第二のデータ記憶領域に書くようになった第二の制御装置と
を有し、
（３）　前記第三の記憶システムは、
前記第一のデータ記憶領域内のデータの複製が格納される第三のデータ記憶領域と、
前記ジャーナルが格納される第三のジャーナル記憶領域と、
独自にスケジュールされたジャーナルリードのタイミングで前記第一の記憶システムから前記ジャーナルを読み、読まれたジャーナルを前記第三のジャーナル記憶領域に書き、そして、独自にスケジュールされたリストアのタイミングで前記第三のジャーナル記憶領域内の前記ジャーナルに基づいて前記第一のデータ記憶領域内のデータの複製を生成して前記第三のデータ記憶領域に書くようになった第三の制御装置と
を有し、
前記第一の記憶システムの前記第一の制御装置は、前記第一のジャーナル記憶領域内の前記ジャーナルが前記第二及び第三の記憶システムによって読まれたか否かを検出し、前記第二及び第三の記憶システムの双方によって読まれるまでは前記第一のジャーナル記憶領域内の前記ジャーナルを保持し、前記第二及び第三の記憶システムの双方によって読まれた後に前記第一のジャーナル記憶領域内の前記ジャーナルを消去可能とするようになっている、
ことを特徴とするデータ処理システム。
【請求項２】
請求項１に記載のデータ処理システムにおいて、
（１）　前記第一の記憶システムは、複数の物理的記憶装置を有し、
前記第一の記憶システムの前記第一の制御装置は、前記上位装置との間でデータを送受する上位アダプタと、前記複数の物理的記憶装置との間でデータを送受するディスクアダプタと、前記上位アダプタで受けるデータ及び前記ディスクアダプタで受けるデータを記憶するキャッシュメモリとを有しており、
前記第一の制御装置は、前記第一の記憶システム内の前記複数の物理的記憶装置が持つ記憶領域を、前記第一のデータ記憶領域及び前記第一のジャーナル記憶領域に割当て、
（２）　前記第二の記憶システムは、複数の物理的記憶装置を有し、

前記第二の記憶システムの前記第二の制御装置は、前記第一の記憶システムとの間でデータを送受する上位アダプタと、前記複数の物理的記憶装置との間でデータを送受するディスクアダプタと、前記上位アダプタで受けるデータ及び前記ディスクアダプタで受けるデータを記憶するキャッシュメモリとを有しており、

前記第二の制御装置は、前記第二の記憶システム内の前記複数の物理的記憶装置が持つ記憶領域を、前記第二のデータ記憶領域及び前記第二のジャーナル記憶領域に割当て、

（２）　前記第三の記憶システムは、複数の物理的記憶装置を有し、

前記第三の記憶システムの前記第三の制御装置は、前記第一の記憶システムとの間でデータを送受する上位アダプタと、前記複数の物理的記憶装置との間でデータを送受するディスクアダプタと、前記上位アダプタで受けるデータ及び前記ディスクアダプタで受けるデータを記憶するキャッシュメモリとを有しており、

前記第三の制御装置は、前記第三の記憶システム内の前記複数の物理的記憶装置が持つ記憶領域を、前記第三のデータ記憶領域及び前記第三のジャーナル記憶領域に割当てるようになっている、

ことを特徴とするデータ処理システム。

【請求項３】

請求項１に記載のデータ処理システムにおいて、

前記第三の記憶システムの前記第三の制御装置は、前記第一のジャーナル記憶領域から読んだジャーナルのデータの数に応じて、前記ジャーナルリードの時間間隔を制御することを特徴とするデータ処理システム。

【請求項４】

請求項１に記載のデータ処理システムにおいて、

前記第三の記憶システムの前記第三の制御装置は、前記第一の記憶システムと前記第三の記憶システムとの間で送受されるデータの通信量に応じて、前記ジャーナルリードの時間間隔を制御するデータ処理システム。

【請求項５】

請求項１記載のデータ処理システムにおいて、

前記第三の記憶システムの前記第三の制御装置は、前記第三のデータ記憶領域が保持している前記ジャーナルの記憶容量に応じて、前記ジャーナルリードの時間間隔を制御するデータ処理システム。

【請求項６】

請求項１に記載のデータ処理システムにおいて、

前記第三の記憶システムの前記第三の制御装置は、前記第三の記憶システムの処理負荷に応じて、前記ジャーナルリードの時間間隔を制御するデータ処理システム。

【請求項７】

請求項１に記載のデータ処理システムにおいて、

前記第三の記憶システムの前記第三の制御装置は、前記第一の記憶システム内の前記第一のジャーナル記憶領域が保持している前記ジャーナルの記憶容量についての情報を、前記第一の記憶システムから読み出し、前記読み出された前記ジャーナルの記憶容量についての情報に応じて、前記ジャーナルリードの時間間隔を制御することを特徴とするデータ処理システム。

【請求項８】

請求項１に記載のデータ処理システムにおいて、

前記第一の記憶システムは、前記第一のジャーナル記憶領域についての管理情報を所有し、

前記第三の記憶システムの前記三の制御装置は、前記第一の記憶システムが所有している前記第一のジャーナル記憶領域についての管理情報を、前記第一の記憶システムから読み出し、読み出された前記第一のジャーナル記憶領域についての管理情報に応じて、前記ジャーナルリードの時間間隔を制御することを特徴とするデータ処理システム。

【請求項９】

請求項１に記載のデータ処理システムにおいて、
前記第一の記憶システム内の前記第一のデータ記憶領域は、複数の論理ボリュームを有し、
前記第一の制御装置は、前記複数の論理ボリュームに格納される複数のデータにそれぞれ対応する複数のジャーナルを、前記第一のジャーナル記憶領域に書き、
前記第一のジャーナル記憶領域に格納される前記複数のジャーナルには、前記複数のジャーナルがそれぞれ対応する前記複数のデータの更新順序に関する情報が含まれており、
前記第二及び第三の記憶システムの前記第二及び第三の制御装置は、それぞれ、前記第一の記憶システムから読んだ前記複数のジャーナルに含まれている前記更新順序に従って、前記複数のジャーナルに基づいて前記複数のデータの複製を生成して前記第二及び第三のデータ記憶領域にそれぞれ書くことを特徴とするデータ処理システム。
【請求項１０】
請求項１に記載のデータ処理システムにおいて、
前記第三の記憶システムは、前記第三の記憶システムの処理負荷に応じて、前記リストアのタイミングを制御することを特徴とするデータ処理システム。
【請求項１１】
上位装置に通信可能に接続された第一の記憶システムと、
前記第一の記憶システムにそれぞれ通信可能に接続された第二及び第三の記憶システムと備え、
（１）　前記第一の記憶システムは、
前記上位装置から送られたデータが格納される第一のデータ記憶領域と、
前記第一の記憶領域に格納されたデータの複製を生成するために使用されるジャーナルが格納される第一のジャーナル記憶領域と、
前記上位装置から送られたデータを前記第一のデータ記憶領域に書き、そして、書かれたデータのジャーナルを前記第一のジャーナル記憶領域に書くようになった第一の制御装置とを有し、
（２）　前記第二の記憶システムは、
前記ジャーナルが格納される第二のジャーナル記憶領域と、
所定のジャーナルリードのタイミングで前記第一の記憶システムから前記ジャーナルを読み、読まれたジャーナルを前記第二のジャーナル記憶領域に書くようになった第二の制御装置とを有し、
（３）　前記第三の記憶システムは、
前記ジャーナルが格納される第三のジャーナル記憶領域と、
所定のジャーナルリードのタイミングで前記第一の記憶システムから前記ジャーナルを読み、読まれたジャーナルを前記第三のジャーナル記憶領域に書くようになった第三の制御装置とを有する、
ことを特徴とするデータ処理システム。
【請求項１２】
請求項１１記載のデータ処理システムにおいて、
前記第二の記憶システムは、前記データの複製を格納する第二のデータ記憶領域を更に有し、前記第二の制御装置が、所定のリストアのタイミングで前記第二のジャーナル記憶領域に格納されたジャーナルから前記データの複製を生成し、生成された前記データの複製を前記第二のデータ記憶装置に書くようになっており、
前記第三の記憶システムは、前記データの複製を格納する第三のデータ記憶領域を更に有し、前記第三の制御装置が、所定のリストアのタイミングで前記第三のジャーナル記憶領域に格納されたジャーナルから前記データの複製を生成し、生成された前記データの複製を前記第三のデータ記憶装置に書くようになっている、
ことを特徴とするデータ処理システム。
【請求項１３】
請求項１１記載のデータ処理システムにおいて、

前記第一の記憶システムの前記第一の制御装置は、前記第一のジャーナル記憶領域内の前記ジャーナルが前記第二及び第三の記憶システムによって読まれたか否かを検出し、前記第二及び第三の記憶システムの双方によって読まれるまでは前記第一のジャーナル記憶領域内の前記ジャーナルを保持し、前記第二及び第三の記憶システムの双方によって読まれた後に前記第一のジャーナル記憶領域内の前記ジャーナルを消去可能とするようになっている、
ことを特徴とするデータ処理システム。

【請求項１４】
請求項１１に記載のデータ処理システムにおいて、
前記第三の記憶システムの前記第三の制御装置は、前記第一のジャーナル記憶領域から読んだジャーナルのデータの数に応じて、前記ジャーナルリードの時間間隔を制御することを特徴とするデータ処理システム。

【請求項１５】
請求項１１に記載のデータ処理システムにおいて、
前記第三の記憶システムの前記第三の制御装置は、前記第一の記憶システムと前記第三の記憶システムとの間で送受されるデータの通信量に応じて、前記ジャーナルリードの時間間隔を制御するデータ処理システム。

【請求項１６】
請求項１１記載のデータ処理システムにおいて、
前記第三の記憶システムの前記第三の制御装置は、前記第三のデータ記憶領域が保持している前記ジャーナルの記憶容量に応じて、前記ジャーナルリードの時間間隔を制御するデータ処理システム。

【請求項１７】
請求項１１に記載のデータ処理システムにおいて、
前記第三の記憶システムの前記第三の制御装置は、前記第三の記憶システムの処理負荷に応じて、前記ジャーナルリードの時間間隔を制御するデータ処理システム。

【請求項１８】
請求項１１に記載のデータ処理システムにおいて、
前記第三の記憶システムの前記第三の制御装置は、前記第一の記憶システム内の前記第一のジャーナル記憶領域が保持している前記ジャーナルの記憶容量についての情報を、前記第一の記憶システムから読み出し、前記読み出された前記ジャーナルの記憶容量についての情報に応じて、前記ジャーナルリードの時間間隔を制御することを特徴とするデータ処理システム。

【請求項１９】
請求項１１に記載のデータ処理システムにおいて、
前記第一の記憶システムは、前記第一のジャーナル記憶領域についての管理情報を所有し、
前記第三の記憶システムの前記三の制御装置は、前記第一の記憶システムが所有している前記第一のジャーナル記憶領域についての管理情報を、前記第一の記憶システムから読み出し、読み出された前記第一のジャーナル記憶領域についての管理情報に応じて、前記ジャーナルリードの時間間隔を制御することを特徴とするデータ処理システム。

【書類名】明細書
【発明の名称】複数の記憶システムを有するデータ処理システム
【技術分野】
【０００１】
本発明は、複数の記憶システムを有するデータ処理システムに関し、特に複数の記憶システム間でのデータの複製に関する。
【背景技術】
【０００２】
近年、常に顧客に対して継続したサービスを提供するために、第一の記憶システムに障害が発生した場合でもデータ処理システムがサービスを提供できるよう、記憶システム間でのデータの複製に関する技術が重要になっている。第一の記憶システムに格納された情報を第二および第三の記憶システムに複製する技術として、以下の特許文献に開示された技術が存在する。
【０００３】
米国特許５１７０４８０号公報には、第一の記憶システムに接続された第一の計算機が、第一の記憶システムに格納されたデータを、第一の計算機と第二の計算機間の通信リンクを介し、第二の計算機に転送し、第二の計算機が第二の計算機に接続された第二の記憶システムに転送する技術が開示されている。
【０００４】
米国特許６２０９００２号公報には、第一の記憶システムが、第一の記憶システムに格納されたデータを、第二の記憶システムに転送し、さらに、第二の記憶システムが、第三の記憶システムに転送する技術が開示されている。計算機と第一の記憶システムとは通信リンクにより接続され、第一の記憶システムと第二の記憶システムとは通信リンクにより接続され、さらに、第二の記憶システムと第三の記憶システムとは通信リンクにより接続されている。第一の記憶システムは、複製対象の第一の論理ボリュームを保持する。第二の記憶システムは、第一の論理ボリュームの複製である第二の論理ボリューム、及び、第二の論理ボリュームの複製である第三の論理ボリュームを保持する。第三の記憶システムは、第三の論理ボリュームの複製である第四の論理ボリュームを保持する。この特許文献において、第二の記憶システムは、第二の論理ボリュームから第三の論理ボリュームへのデータ複製処理と、第三の論理ボリュームから第四の論理ボリュームへのデータ複製処理とを排他的に実行する。
【０００５】
【特許文献１】米国特許５１７０４８０号公報
【特許文献２】米国特許６２０９００２号公報
【発明の開示】
【発明が解決しようとする課題】
【０００６】
米国特許５１７０４８０号公報に開示された技術は、データの複製のために、第一の計算機および第二の計算機を常に使用する。第一の計算機は、通常業務を行っており、第一の計算機にかかるデータ複製処理の負荷は無視できない。さらに、複製のためのデータは、第一の計算機と第一の記憶システム間の通信リンクを使用するため、通常業務のために必要なデータ転送と衝突し、通常業務に必要なデータ参照、データ更新時間が長くなるという課題がある。
【０００７】
米国特許６２０９００２号公報に開示された技術は、複製を行うデータ量の倍の記憶容量が第二の記憶システム及び第三の記憶システムに必要となる。また、複製対象のデータ量が多いことにより、データ複製処理に費やされる時間が長く、第三の記憶システムのデータは古いものとなってしまう。その結果、第三の記憶システムのデータを用いて業務が再開される場合、第三の記憶システムのデータを最新とするまでの時間が長くなり、業務再開までの時間が延びるという課題がある。さらに、この文献の開示によれば、第一の記

憶システムは、第一の記憶システム内のデータ更新処理に加えて、第二の記憶システムとの間でのデータ更新処理が終了した時点で、上位の計算機にデータ更新完了報告を行う。したがって、計算機からのデータ更新に費やされる時間が長く、第一の記憶システムと第二の記憶システムとの間の距離が遠くなればなるほど、データ更新に費やされる時間はより長くなる。その結果、各記憶システム間の距離をあまり遠くすることができないという問題もある。

【０００８】

本発明の目的は、記憶システムの上位の計算機に影響を与えず、複数の記憶システム間でデータ転送又はデータの複製を行えるようにすることにある。

【０００９】

本発明の別の目的は、記憶システムと計算機との間の通信に影響を与えずに、複数の記憶システム間でデータ転送又はデータの複製をすることにある。

【００１０】

本発明のまた別の目的は、複数の記憶システム内に保持されるデータ格納領域を少なくすることにある。

【００１１】

本発明のさらに別の目的は、複数の記憶システムの上位の計算機の業務に影響を与えることのないように、高速かつ効率的に複数の記憶システム間でデータ転送又はデータの複製をすることにある。

【課題を解決するための手段】

【００１２】

本発明の一つの観点に従うデータ処理システムは、上位装置に通信可能に接続された第一の記憶システムと、第一の記憶システムにそれぞれ通信可能に接続された第二及び第三の記憶システムと備える。

【００１３】

第一の記憶システムは、第一の制御装置と、上位装置から送られたデータが格納される第一のデータ記憶領域と、第一の記憶領域に格納されたデータの複製を生成するために使用されるジャーナルが格納される第一のジャーナル記憶領域とを有する。第一の制御装置は、上位装置から送られたデータを第一のデータ記憶領域に書き、第一のデータ記憶領域に書かれたデータのジャーナルを第一のジャーナル記憶領域に書き、そして、第二及び第三の記憶システムのそれぞれからの要求に応じて、第一のジャーナル記憶領域内のジャーナルを第二及び第三の記憶システムにそれぞれ送るようになっている。

【００１４】

第二の記憶システムは、第二の制御装置と、第一のデータ記憶領域内のデータの複製が格納される第二のデータ記憶領域と、ジャーナルが格納される第二のジャーナル記憶領域とを有する。第二の制御装置は、独自にスケジュールされたジャーナルリードのタイミングで第一の記憶システムからジャーナルを読み、読まれたジャーナルを第二のジャーナル記憶領域に書き、そして、独自にスケジュールされたリストアのタイミングで、第二のジャーナル記憶領域内の前記ジャーナルに基づいて第一のデータ記憶領域内のデータの複製を生成して、これを第二のデータ記憶領域に書くようになっている。

【００１５】

三の記憶システムは、第三の制御装置と、第一のデータ記憶領域内のデータの複製が格納される第三のデータ記憶領域と、ジャーナルが格納される第三のジャーナル記憶領域とを有する。第三の制御装置は、独自にスケジュールされたジャーナルリードのタイミングで第一の記憶システムからジャーナルを読み、読まれたジャーナルを第三のジャーナル記憶領域に書き、そして、独自にスケジュールされたリストアのタイミングで第三のジャーナル記憶領域内のジャーナルに基づいて第一のデータ記憶領域内のデータの複製を生成して、これを第三のデータ記憶領域に書くようになっている。

【００１６】

さらに、第一の記憶システム内の第一の制御装置は、第一のジャーナル記憶領域内のジ

ャーナルが第二及び第三の記憶システムによって読まれたか否かを検出する。そして、第一の制御装置は、第二及び第三の記憶システムの双方によって読まれるまでは、第一のジャーナル記憶領域内のジャーナルを保持し、第二及び第三の記憶システムの双方によって読まれた後に、第一のジャーナル記憶領域内のジャーナルを消去可能とするようになっている。

【００１７】

第三の記憶システムの第三の制御装置は、第一のジャーナル記憶領域から読んだジャーナルのデータの数に応じて、ジャーナルリードの時間間隔を制御するようになっていてよい。或いは、第三の制御装置は、第一の記憶システムと第三の記憶システムとの間で送受されるデータの通信量に応じて、ジャーナルリードの時間間隔を制御するようになっていてもよい。或いは、第三の制御装置は、第三のデータ記憶領域が保持しているジャーナルの記憶容量に応じて、ジャーナルリードの時間間隔を制御するようになっていてもよい。或いは、第三の制御装置は、第三の記憶システムの処理負荷に応じて、ジャーナルリードの時間間隔を制御するようになっていてもよい。或いは、第三の制御装置は、第一の記憶システム内の第一のジャーナル記憶領域が保持しているジャーナルの記憶容量についての情報を、第一の記憶システムから読み出し、読み出されたジャーナルの記憶容量についての情報に応じて、ジャーナルリードの時間間隔を制御するようになっていてもよい。或いは、第一の記憶システムが、第一のジャーナル記憶領域についての管理情報を所有しており、そして、第三の記憶システムの第三の制御装置が、前記第一の記憶システムが所有している前記第一のジャーナル記憶領域についての管理情報を、前記第一の記憶システムから読み出し、読み出された第一のジャーナル記憶領域についての管理情報に応じて、ジャーナルリードの時間間隔を制御するようにしてもよい。第二の記憶システムの第二の制御装置も、基本的に、これと同様にジャーナルリードの時間間隔を制御するようになっていてよい。

【００１８】

第一の記憶システム内の第一のデータ記憶領域は、複数の論理ボリュームから構成されることができる。第一の制御装置は、その複数の論理ボリュームに格納される複数のデータにそれぞれ対応する複数のジャーナルを、第一のジャーナル記憶領域に書くことができる。第一のジャーナル記憶領域に格納される複数のジャーナルには、対応する複数のデータの更新順序に関する情報が含まれることができる。第二及び第三の記憶システムの第二及び第三の制御装置は、それぞれ、第一の記憶システムから読んだ複数のジャーナルに含まれている更新順序に従って、複数のジャーナルに基づいて複数のデータの複製を生成して、それを第二及び第三のデータ記憶領域にそれぞれ書くことができる。

【００１９】

第三の記憶システムは、第三の記憶システムの処理負荷に応じて、リストアのタイミングを制御するようにしてもよい。

【００２０】

本発明の別の観点に従う上位装置に通信可能に接続された第一の記憶システムと、第一の記憶システムにそれぞれ通信可能に接続された第二及び第三の記憶システムと備える。

【００２１】

第一の記憶システムは、第一の制御装置と、上位装置から送られたデータが格納される第一のデータ記憶領域と、第一の記憶領域に格納されたデータの複製を生成するために使用されるジャーナルが格納される第一のジャーナル記憶領域とを有する。この第一の記憶システムの第一の制御装置は、上位装置から送られたデータを第一のデータ記憶領域に書き、そして、書かれたデータのジャーナルを第一のジャーナル記憶領域に書くようになっている。

【００２２】

第二の記憶システムは、第二の制御装置と、ジャーナルが格納される第二のジャーナル記憶領域とを有する。この第二の記憶システムの第二の制御装置は、所定のジャーナルリードのタイミングで第一の記憶システムからジャーナルを読み、読まれたジャーナルを第

二のジャーナル記憶領域に書くようになっている。

【００２３】

第三の記憶システムは、第三の制御装置と、ジャーナルが格納される第三のジャーナル記憶領域とを有する。この第三の記憶システムの第三の制御装置は、所定のジャーナルリードのタイミングで第一の記憶システムからジャーナルを読み、読まれたジャーナルを第三のジャーナル記憶領域に書くようになっている。

【００２４】

第二の記憶システムは、データの複製を格納する第二のデータ記憶領域を更に有することができる。第二の制御装置は、所定のリストアのタイミングで第二のジャーナル記憶領域に格納されたジャーナルからデータの複製を生成し、生成されたデータの複製を第二のデータ記憶装置に書くようになっていてよい。

【００２５】

第三の記憶システムも、データの複製を格納する第三のデータ記憶領域を更に有することができる。第三の制御装置は、所定のリストアのタイミングで第三のジャーナル記憶領域に格納されたジャーナルからデータの複製を生成し、生成されたデータの複製を第三のデータ記憶装置に書くようになっていてよい。

【００２６】

第一の記憶システムの第一の制御装置は、第一のジャーナル記憶領域内のジャーナルが第二及び第三の記憶システムによって読まれたか否かを検出し、第二及び第三の記憶システムの双方によって読まれるまでは第一のジャーナル記憶領域内の前記ジャーナルを保持し、第二及び第三の記憶システムの双方によって読まれた後に第一のジャーナル記憶領域内のジャーナルを消去可能とするようになっていてよい。

【発明を実施するための最良の形態】

【００２７】

本発明によるデータ処理システムの実施形態を図面により詳細に説明する。

【００２８】

図１は、本発明に従うデータ処理システムの一実施形態の物理的な構成を示すブロック図である。図２は、本実施形態の論理的な構成を示すブロック図である。

【００２９】

図１及び図２に示すように、このデータ処理システムでは、ホストコンピュータ１８０と記憶システム１００Ａが接続パス１９０により接続され、また、記憶システム１００Ａと別の記憶システム１００Ｂとが接続パス２００により接続される。記憶システム１００Ｂは、記憶システム１００Ａに保存されたデータの複製を保持するために使用される。以下の説明において、複製対象のデータを保持する記憶システム１００Ａと複製データを保持する記憶システム１００Ｂとの区別を容易とするために、前者１００Ａを「正記憶システム」、後者１００Ｂを「副記憶システム」と呼ぶこととする。

【００３０】

正記憶システム１００Ａと副記憶システム１００Ｂの物理的構成について、図１を参照して説明する。

【００３１】

正記憶システム１００Ａと副記憶システム１００Ｂは基本的には同様の物理的構成を有するので、図１では、代表的に正記憶システム１００Ａの構成のみが示されている。図１に示すように、正記憶システム１００Ａは、１つ以上のチャネルアダプタ１１０、１つ以上のディスクアダプタ１２０、１つ以上のキャッシュメモリ１３０、１つ以上の共有メモリ１４０、１つ以上の物理的記憶装置（例えば、ハードディスクドライブ）１５０、１つ以上のコモンパス１６０、１つ以上の接続線１７０を備える。チャネルアダプタ１１０、ディスクアダプタ１２０、キャッシュメモリ１３０、共有メモリ１４０はコモンパス１６０により相互に接続されている。コモンパス１６０は、コモンパス１６０の障害時のために２重化されてもよい。ディスクアダプタ１２０と物理的記憶装置１５０とは接続線１７０によって接続されている。また、図示していないが、記憶システム１００の設定、監視

、保守等を行うための保守端末が全てのチャネルアダプタ１１０とディスクアダプタ１２０とに専用線を用いて接続されている。

【００３２】

チャネルアダプタ１１０は、接続線１９０又は接続線２００によりホストコンピュータ１８０又は他の記憶システム（例えば副記憶システム１００Ｂ）と接続される。チャネルアダプタ１１０は、ホストコンピュータ１８０とキャッシュメモリ１３０間のデータ転送、又は他の記憶システムとキャッシュメモリ１３０間のデータ転送を制御する。ディスクアダプタ１２０は、キャッシュメモリ１３０と物理的記憶装置１５０との間のデータ転送を制御する。キャッシュメモリ１３０は、ホストコンピュータ１８０又は他の記憶システムから受信したデータ、あるいは物理的記憶装置１５０から読み出したデータを一時的に保持するメモリである。共有メモリ１４０は、記憶システム１００内の全てのチャネルアダプタ１１０とディスクアダプタ１２０とが共有するメモリである。共有メモリ１４０には、主に、チャネルアダプタ１１０とディスクアダプタ１２が使用する制御や管理のための様々な情報（例えば、後述するボリューム情報４００、ペア情報５００、グループ情報６００およびポインタ情報７００など）が記憶され保持される。副記憶システム１００Ｂの物理的構成も、基本的にこれと同様である。

【００３３】

次に、記憶システム１００Ａ及び１００Ｂの論理的な構成について、図２を参照して説明する。

【００３４】

図２に示すように、記憶システム１００Ａ及び１００Ｂの各々において、物理的記憶装置１５０、１５０、…により提供される全体の記憶領域は、論理的な多数の記憶領域２３０、２３０、…に分割されて管理される。以下、個々の論理的な記憶領域２３０を「論理ボリューム」と呼ぶこととする。各論理ボリューム２３０の容量および記憶システム１００Ａ又は１００Ｂ内での物理的な格納位置（物理アドレス）は、記憶システム１００Ａ又は１００Ｂに接続された保守用の端末コンピュータ（図示せず）もしくはホストコンピュータ１８０から指定することができる。各論理ボリューム２３０の物理アドレスは、後述するボリューム情報４００に保存される。物理アドレスは、例えば、記憶システム１００Ａ又は１００Ｂ内の物理的記憶装置１５０を識別する番号（記憶装置番号）と、その記憶装置１５０内での記憶領域を一意に示す数値（例えば、物理的記憶装置１５０内での記憶領域の先頭からの位置）から構成される。以下の説明では、物理アドレスは、物理的記憶装置１５０の記憶装置番号とその物理的記憶装置１５０内での記憶領域の先頭からの位置の組で表されるものとする。普及しているＲＡＩＤの原理に従った記憶システムでは、１つの論理ボリューム２３０が複数の物理的記憶装置１５０内の複数の物理的な記憶領域に対応しているが、以下では、説明を容易にするために、１つの論理ボリューム２３０が、１つの物理的な記憶装置１５０内の１つの記憶領域に対応するものとして説明する。しかし、ＲＡＩＤの原理に従う記憶システムでも本発明の原理が以下の説明と同様に適用できることを、当業者は容易に理解する筈である。

【００３５】

記憶システム１００Ａ、１００Ｂ内に保存されているデータは、そのデータが存在する論理ボリューム２３０を識別する番号（論理ボリューム番号）と、その論理ボリューム２０３０内のそのデータの記憶領域を一意に示す数値（例えば、論理ボリューム２３０の記憶領域の先頭からの位置）により一意に指定することができる。以下の説明では、論理アドレスは、論理ボリューム２３０の論理ボリューム番号と論理ボリューム２３０内での記憶領域の先頭からの位置（論理ボリューム内位置）の組で表されるものとする。

【００３６】

以下の説明において、複製対象の論理ボリュームとその複製である論理ボリュームとの区別を容易とするために、前者を「正論理ボリューム」、後者を「副論理ボリューム」とよぶこととする。一対の正論理ボリュームと副論理ボリュームを「ペア」とよぶこととする。正論理ボリュームと副論理ボリュームの関係および状態等の情報は後述するペア情報

５００に保存される。

【００３７】

複数の正論理ボリュームと、それらの正論理ボリュームとペアをそれぞれ構成する複数の副論理ボリュームとの間のデータの更新順序を一致させるために、「グループ」という管理単位が設けられる。例えば、ホストコンピュータ１８０が、第１の正論理ボリューム内の第１のデータを更新し、その後、その第１のデータを読み出し、その第１のデータを用いて、第２の正論理ボリューム内の第２のデータを更新する処理を行う場合を想定する。この場合、第１の正論理ボリュームから第１の副論理ボリュームへのデータ複製処理と、第２の正論理ボリュームから第２の副論理ボリュームへのデータ複製処理とが独立に行われたならば、第１の副論理ボリュームへの第１のデータの複製処理より前に、第２の副論理ボリュームへの第２のデータの複製処理が行われる場合がある。この場合に、もし、第２の副論理ボリュームへの第２のデータの複製処理の終了後、第１の副論理ボリュームへの第１のデータの複製処理が終了する前に、故障等が発生して、第１の副論理ボリュームへの複製処理が停止したならば、第1の副論理ボリュームと第２の副論理ボリュームの間でのデータの整合性が損なわれることになる。このような場合であって、第１の副論理ボリュームと第２の副論理ボリューム間のデータの整合性を保つために、「グループ」を用いた更新順序の制御が行われる。すなわち、正論理ボリュームと副論理ボリューム間でデータの更新順序を一致させる必要のある複数の論理ボリュームが、同じグループに登録される。そのグループに属する正論理ボリュームで行われるデータの更新毎に、後述するグループ情報６００の更新番号が割り当てられる。そして、その更新番号順に、更新されたデータの副論理ボリュームへの複製処理が行われる。例えば、図２の例では、正記憶システム１００Ａ内の２つの正論理ボリューム「ＤＡＴＡ１」と「ＤＡＴＡ２」が一つのグループ「グループ１」を構成する。そして、これら２つの正論理ボリューム「ＤＡＴＡ１」と「ＤＡＴＡ２」の複製である２つの副論理ボリューム「ＣＯＰＹ１」と「ＣＯＰＹ２」が、副記憶システム１００Ｂ内で同じグループ「グループ１」を構成する。

【００３８】

データ複製対象である正論理ボリュームのデータを更新する場合、副論理ボリュームのデータの更新に利用するために、そのデータ更新についてのジャーナルが作成され、そのジャーナルが正記憶システム１００Ａ内の所定の論理ボリュームに保存される。本実施形態では、正記憶システム１１０Ａ内の各グループに対して、そのグループのジャーナルのみを保存するための論理ボリューム（以下、ジャーナル論理ボリュームとよぶ）が割り当てられる。図２の例では、グループ「グループ１」に、ジャーナル論理ボリューム「ＪＮＬ１」が割り当てられている。副記憶システム１００Ｂ内の各グループにも、ジャーナル論理ボリュームが割り当てられる。図１では、副記憶システム内のグループ「グループ１」に、ジャーナル論理ボリューム「ＪＮＬ２」が割り当てられている。副記憶システム１００Ｂ内のジャーナル論理ボリューム「ＪＮＬ２」は、正記憶システム１００Ａから副記憶システム１００Ｂに転送された「グループ１」のジャーナルを保存するために使用される。ジャーナルを副記憶システム１００Ｂ内のジャーナル論理ボリュームに保存することにより、ジャーナル受信時に副論理ボリュームのデータ更新を行なう必要がなくなり、ジャーナル受信時とは非同期の後の時期に、例えば副記憶システム１００Ｂの負荷が低い時に、そのジャーナルに基づいて副論理ボリュームのデータを更新することができる。

【００３９】

さらに、正記憶システム１００Ａと副記憶システム１００Ｂとの間の接続線２００が複数ある場合、正記憶システム１００Ａから副記憶システム１００Ｂへのジャーナルの転送を多重に行い、複数の接続線２００の転送能力を有効に利用することができる。所定の更新順序を守ることにより、副記憶システム１００Ｂに多くのジャーナルが溜まる可能性があるが、副論理ボリュームのデータ更新に直ぐに使用できないジャーナルは、ジャーナル論理ボリュームに退避することにより、キャッシュメモリを開放することができる。

【００４０】

上述のジャーナルは、ライトデータと更新情報とから構成される。更新情報は、ライト

データを管理するための情報で、ライト命令を受信した時刻、グループ番号、後述するグループ情報６００の更新番号、ライト命令の論理アドレス、ライトデータのデータサイズ、ライトデータを格納したジャーナル論理ボリュームの論理アドレス等を含む。更新情報は、ライト命令を受信した時刻と更新番号のどちらか一方のみを保持してもよい。ホストコンピュータ１８０からのライト命令の中にライト命令の作成時刻が含まれている場合は、ライト命令を受信した時刻の代わりに、当該ライト命令内の命令作成時刻を使用してもよい。

【００４１】

図３はジャーナルの更新情報とライトデータの関係の一例を示す。

【００４２】

図３に示すように、ジャーナル論理ボリューム３５０は、例えば、更新情報を格納する記憶領域（更新情報領域）３５１と、ライトデータを格納する記憶領域（ライトデータ領域）３５３に分割されて使用される。正論理ボリューム３４に書き込まれたライトデータ３２０の更新情報３１０は、ジャーナル論理ボリューム３５０の更新情報領域３５１の先頭から、更新番号の順に格納される。或る更新情報３１０の格納場所が更新情報領域３５１の終端に達すると、後続の更新情報３１０は更新情報領域３５１の先頭から格納される。正論理ボリューム３４に書き込まれたライトデータ３２０に対応するライトデータ３３０が、ジャーナル論理ボリューム３５０のライトデータ領域３５３先頭から、更新番号の順に格納される。或るライトデータ３３０の格納場所がライトデータ領域３５３の終端に達すると、後続のライトデータ３３０はライトデータ領域３５１の先頭から格納される。更新情報領域３５１およびライトデータ領域３５３のサイズ比は、固定値でもよいし、保守端末あるいはホストコンピュータ１８０により任意値に設定可能であってもよい。ジャーナル論理ボリューム３５０内での更新情報領域３５１およびライトデータ領域３５３の範囲を示したアドレスや、最新及び最古の更新情報３１０およびライトデータ３３０の記憶場所を示したアドレスなどの情報は、後述するポインタ情報７００内に保持される。以下の説明では、ジャーナル論理ボリューム３５０が更新情報領域３５１とライトデータ領域３５３に分割されて使用されるが、変形例として、ジャーナル論理ボリューム３５０の先頭から、各ジャーナルの更新情報３１０とライトデータ３３０を連続的に格納される方式を採用してもよい。

【００４３】

図４は、ジャーナルの更新情報３１０の具体例を示す。

【００４４】

図４に例示された更新情報３１０には、「１９９９年３月１７日の２２時２０分１０秒」のようなライト命令の受信時刻が記憶されている。この更新情報３１０には、また、当該ライト命令が、ライトデータ３２０を論理ボリューム番号「１」の論理ボリュームの記憶領域の先頭から「７００」の位置に格納する命令であったこと、および、ライトデータのサイズは「３００」であったことが記録されている。さらに、この更新情報３１０には、ジャーナルのライトデータ３１０が、論理ボリューム番号「４」（ジャーナル論理ボリューム）の記憶領域の先頭から「１５００」の位置から格納されていることが記録されている。また、この更新情報３１０には、論理ボリューム番号「１」の論理ボリュームはグループ「１」に属すること、および、上記ライト命令によるデータ更新が、グループ「１」のデータ複製開始から「４」番目のデータ更新であることが記録されている。

【００４５】

再び図２を参照する。正記憶システム１００Ａは、ホストコンピュータ１８０からのデータのリード／ライト命令を受付ける命令受付処理２１０、及びデータを然るべき論理ボリューム２３０にリード／ライトするリードライト処理２２０を行う機能を有する。副記憶システム１００Ｂは、正記憶システム１００Ａからジャーナルを読むジャーナルリード（ＪＮＬＲＤ）処理２４０、データを然るべき論理ボリューム２３０にリード／ライトするリードライト処理２２０、及び正記憶システム１００Ａからのジャーナルに基づいて然るべき副論理ボリューム２３０内のデータを更新するリストア処理２５０を行う機能を有

する。これらの処理機能は、図２に示したチャネルアダプタ１１０、ディスクアダプタ１２０、キャッシュメモリ１３０及び共有メモリ１４０によって実現される。

【００４６】

図２を参照して、正記憶システム１００Ａの正論理ボリュームへのデータ更新を副記憶システム１００Ｂの副論理ボリュームに反映する動作について概説する。

【００４７】

（１）　正記憶システム１００Ａは、ホストコンピュータ１８０から正論理ボリューム（例えば「ＤＡＴＡ１」）２３０内のデータに対するライト命令を受信すると、命令受信処理２１０およびリードライト処理２２０によって、正論理ボリューム（「ＤＡＴＡ１」）２３０内の対象データの更新と、ジャーナル論理ボリューム（「ＪＮＬ１」）２３０へのジャーナルの保存とを行う(矢印２７０)。

【００４８】

（２）　副記憶システム１００Ｂは、ジャーナルリード処理２４０によって、正記憶システム１００Ａにジャーナルをリードする命令を送る。正記憶システム１００Ａは、副記憶システム１００Ｂからジャーナルをリードする命令を受信すると、命令受信処理２１０およびリードライト処理２２０によって、ジャーナル論理ボリューム（「ＪＮＬ１」）２３０からジャーナルを読み出し、副記憶システム１００Ｂに送信する(矢印２８０)。

【００４９】

（３）　副記憶システム１００Ｂは、正記憶システムからリードしたジャーナルを、リードライト処理２２０によって、ジャーナル論理ボリューム（「ＪＮＬ２」）２３０に保存する(矢印２８０)。

【００５０】

（４）　副記憶システム１００Ｂは、リストア処理２５０およびリードライト処理２２０によって、ポインタ情報７００を用いて、更新番号の昇順に、ジャーナル論理ボリューム（「ＪＮＬ２」）２３０からジャーナルを読み出し、副論理ボリューム（例えば「ＣＯＰＹ１」）２３０のデータを更新する(矢印２９０)。

【００５１】

図５は、ボリューム情報４００の具体例を示す。

【００５２】

ボリューム情報４００は、チャネルアダプタ１１０およびディスクアダプタ１２０から参照可能なメモリ、例えば図１に示した共有メモリ１４０に保存される。ボリューム情報４００は、論理ボリュームを管理するための情報であり、図５に示すように、論理ボリューム番号毎に、ボリューム状態、フォーマット形式、容量、ペア番号、物理アドレスを保持する。ボリューム状態は、“正常”、“正”、“副”、“異常”、“未使用”のいずれかを保持する。ボリューム状態が“正常”もしくは“正”である論理ボリューム２３０は、ホストコンピュータ１８０から正常にアクセス可能な論理ボリューム２３０である。ボリューム状態が“副”である論理ボリューム２３０は、ホストコンピュータ１８０からのアクセスを許可してもよい。ボリューム状態が“正”である論理ボリューム２３０は、データの複製が行われている正論理ボリューム２３０である。ボリューム状態が“副”である論理ボリューム２３０は、複製に使用されている副論理ボリューム２３０である。ボリューム状態が“異常”の論理ボリューム２３０は、障害により正常にアクセスできない論理ボリューム２３０である。ここで、障害とは、例えば、論理ボリューム２３０を保持する記憶装置１５０の故障である。ボリューム状態が“未使用”の論理ボリューム２３０は、使用していない論理ボリューム２３０である。ペア番号は、ボリューム状態が“正”もしくは“副”の場合に有効であり、後述するペア情報５００を特定するためのペア番号を保持する。図５に示す例では、論理ボリューム番号「１」の論理ボリュームは、複製対象のデータを保持した正論理ボリュームであり、アクセス可能であり、フォーマット形式が「ＯＰＥＮ３」であり、容量が「３ＧＢ」であり、そこに保持されたデータは、記憶装置番号「１」の物理的記憶装置１５０の記憶領域の先頭から格納されていることが、示されている。

【００５３】
図６は、ペア情報５００の具体例を示す。
【００５４】
ペア情報５００は、チャネルアダプタ１１０およびディスクアダプタ１２０から参照可能なメモリ、例えば図１に示した共有メモリ１４０に保存される。ペア情報５００は、ペアを管理するための情報であり、図６に示すように、ペア番号毎に、ペア状態、正記憶システム番号、正論理ボリューム番号、副記憶システム番号、副論理ボリューム番号、グループ番号、コピー済みアドレスを保持する。ペア状態は、“正常”、“異常”、“未使用”，“コピー未”、“コピー中”のいずれかを保持する。ペア状態が“正常”の場合は、正論理ボリューム２３０のデータ複製が正常に行われていることを示す。ペア状態が“異常”の場合は、障害により正論理ボリューム２３０の複製が行えないことを示す。ここで、障害とは、例えば、接続パス２００の断線などである。ペア状態が“未使用”の場合は、当該ペア番号の情報は有効でないことを示す。ペア状態が“コピー中”の場合は、後述する初期コピー処理中であることを示す。ペア状態が“コピー未”の場合は、後述する初期コピー処理が未だ行われていないことを示す。正記憶システム番号は、正論理ボリューム２３０を保持する正記憶システム１００Ａを特定する番号を保持する。副記憶システム番号は、副論理ボリューム２３０を保持する副記憶システム１００Ｂを特定する番号を保持する。グループ番号は、正記憶システム１００Ａの場合は、正論理ボリュームが属するグループ番号を保持し、副記憶システム１００Ｂの場合は、副論理ボリュームが属するグループ番号を保持する。コピー済みアドレスは、後述する初期コピー処理にて説明する。図６に例示したペア情報１は、データ複製対象が正記憶システム番号「１」の正論理ボリューム番号「１」により特定される正論理ボリュームであり、データ複製先が副記憶システム番号「２」の副論理ボリューム番号「１」で特定される副論理ボリュームであり、正常にデータ複製処理が行われていることを示す。
【００５５】
図７は、グループ情報６００の具体例を示す。
【００５６】
グループ情報６００は、チャネルアダプタ１１０およびディスクアダプタ１２０から参照可能なメモリ、例えば共有メモリ１４０に保存される。図７に示すように、グループ情報６００は、グループ番号毎に、グループ状態、ペア集合、ジャーナル論理ボリューム番号、更新番号を保持する。グループ状態は、“正常”、“異常”、“未使用”のいずれかを保持する。グループ状態が“正常”の場合は、ペア集合の少なくともひとつのペア状態が“正常”であることを示す。グループ状態が“異常”の場合は、ペア集合の全てのペア状態が“異常”であることを示す。グループ状態が“未使用”の場合は、当該グループ番号の情報は有効でないことを示す。ペア集合は、正記憶システムの場合は、そのグループに属する全ての正論理ボリュームのペア番号を保持し、副記憶システムの場合は、そのグループに属する全ての副論理ボリュームのペア番号を保持する。ジャーナル論理ボリューム番号は、当該グループ番号のグループに属するジャーナル論理ボリューム番号を示す。更新番号は、初期値は１であり、グループ内の正論理ボリュームに対しデータの書き込みが行われると、＋１だけ増加する。更新番号は、ジャーナルの更新情報に記憶され、副記憶システム１００Ｂにて、データの更新順を守るために使用する。図７に零時さえタグループ情報６００は、グループ番号「１」のグループが、ペア番号「１」及び「２」に属する論理ボリュームと、論理ボリューム番号「４」のジャーナル論理ボリューム４から構成されており、正常にデータの複製処理が行われていることを示す。
【００５７】
図８は、ポインタ情報７００の具体例を示す。図９は、ポインタ情報に含まれる項目とジャーナル論理ボリューム３５０との関係を説明した図である。
【００５８】
ポインタ情報７００は、チャネルアダプタ１１０およびディスクアダプタ１２０から参照可能なメモリ、例えば共有メモリ１４０に保存される。ポインタ情報７００は、グルー

プ毎に、当該グループのジャーナル論理ボリュームを管理するための情報であり、図８に示すように、そのジャーナル論理ボリューム内の更新情報領域先頭アドレス、ライトデータ領域先頭アドレス、更新情報最新アドレス、更新情報最古アドレス、ライトデータ最新アドレス、ライトデータ最古アドレス、リード開始アドレス及びリトライ開始アドレスを保持する。

【００５９】

正記憶システム１００Ａは、正記憶システム１００Ａ内のジャーナル論理ボリュームを管理するためのポインタ情報７００を、正記憶システム１００Ａに接続された副記憶システムの台数分のセット数だけ有する。すなわち、図２に示した例では、正記憶システム１００Ａに接続された副記憶システム１００Ｂは１台であるため、図８に示すように、正記憶システム１００Ａは、その一つの副記憶システム１００Ｂに対応した１セットのポインタ情報７００を有し、そのポインタ情報７００には、その一つの副記憶システム１００Ｂのシステム番号が記載されている。しかし、後に説明する図２５の例では、正記憶システム１００Ａに複数(例えば２台）の副記憶システム１００Ｂ、１００Ｃが並列に接続される。図２５の例の場合、正記憶システム１００Ａは、それら複数の副記憶システム１００Ｂ、１００Ｃにそれぞれに対応した複数セットのポインタ情報７００Ｂ、７００Ｃ（図２６参照）を有し、それらの複数セットのポインタ情報７００Ｂ、７００Ｃには対応する副記憶システム１００Ｂ、１００Ｃのシステム番号が記録されている。このように正記憶システム１００Ａが、それに接続された複数の副記憶システム１００Ｂ、１００Ｃにそれぞれ対応する複数のポインタ情報７００を有する場合、それぞれの副記憶システム１００Ｂ、１００Ｃによる正記憶システム１００Ａからのジャーナルリードが行われたか否か（そのタイミングは副記憶システム１００Ｂ、１００Ｃによって異なる）が、副記憶システム１００Ｂ、１００Ｃにそれぞれ割り当てられた複数のポインタ情報７００によって管理することができる。これを利用して、正記憶システム１００Ａは、正記憶システム１００Ａ内の各ジャーナルを、それが複数の副記憶システム１００Ｂ、１００Ｃの全てによってリードされ終わらない限り、消去されないよう保持するよう制御を行うことができる。

【００６０】

一方、副記憶システム１００Ｂは、基本的に、その副記憶システム１１０Ｂ内のジャーナル論理ボリュームを管理するための1セットのポインタ情報７００を有する。ただし、特に図示されてはいないが、正記憶システムに複数の副記憶システムがカスケード接続された構成も採用でき、この場合、そのカスケードの中間に位置する副記憶システムは、その上側と下側の２つの副記憶システムにそれぞれ対応した複数セットのポインタ情報をもつことができる。

【００６１】

図９に示すように、更新情報領域先頭アドレスは、ジャーナル論理ボリューム３５０の更新情報領域３５３の先頭の論理アドレスを保持する。ライトデータ領域先頭アドレスは、ジャーナル論理ボリューム３５０のライトデータ領域３５３も先頭の論理アドレスを保持する。更新情報最新アドレスは、次にジャーナルを格納する場合に、その更新情報の保存に使用される先頭の論理アドレスを保持する。更新情報最古アドレスは、最古の（更新番号が小さい）ジャーナルの更新情報を保存する先頭の論理アドレスを保持する。ライトデータ最新アドレスは、次にジャーナルを格納する場合に、ライトデータの保存に使用する先頭の論理アドレスを保持する。ライトデータ最古アドレスは、ジャーナル論理ボリューム３５０内での最古の（更新番号が最も小さい）ジャーナルのライトデータを保存する先頭の論理アドレスを保持する。リード開始アドレスとリトライ開始アドレスは、正記憶システム１００Ａのみで使用され、後述するジャーナルリード受信処理にて使用される。図８に例示されたポインタ情報７００によれば、更新情報領域３５が論理ボリュー番号「４」の記憶領域の先頭（アドレス「０」）からアドレス「６９９」の位置までであり、ライトデータ領域３５３が論理ボリューム番号「４」の記憶領域のアドレス「７００」の位置からアドレス「２６９９」の位置までである。そして、更新情報は、論理ボリューム番号「４」の記憶領域のアドレス「２００」の位置からアドレス「４９９」の位置まで保存

されており、次のジャーナルの更新情報は、論理ボリューム番号「４」の記憶領域のアドレス「５００」の位置から格納されることになる。また、ジャーナルのライトデータは、論理ボリューム番号「４」の記憶領域のアドレス「１３００」の位置からアドレス「２１９９」の位置までに保存されており、次のジャーナルのライトデータは、論理ボリューム番号「４」の記憶領域のアドレス「２２００」の位置から格納されることになる。

【００６２】

ここでの説明では、１つのグループに１つのジャーナル論理ボリュームが割り当てられる。しかし、１つのグループに複数のジャーナル論理ボリュームを割り当てられてもよい。例えば、１つのグループに２つのジャーナル論理ボリュームが割り当てられ、ジャーナル論理ボリューム毎にポインタ情報７００が設けられ、そして、その２つのジャーナル論理ボリュームに交互にジャーナルが格納されるようにしてもよい。これにより、ジャーナルの物理的記憶装置１５０への書き込みが分散され、性能の向上が見込める。さらに、ジャーナルのリード性能も向上する。別の例としては、１つのグループに２つのジャーナル論理ボリュームが割り当てられ、通常は、１つのジャーナル論理ボリュームのみが使用される。もう一方のジャーナル論理ボリュームは、現在使用されてきたジャーナル論理ボリュームの性能が低下した場合に使用される。性能が低下する場合とは、例えば、ジャーナル論理ボリュームが、複数の物理的記憶装置１５０から構成され、ＲＡＩＤ５の方式でデータを保持しているところ、それら複数の物理的記憶装置１５０の一台が故障した場合である。

【００６３】

上述したように、ボリューム情報４００、ペア情報５００、グループ情報６００、及びポインタ情報7.００等は、共有メモリ１４０に格納されることができる。しかし、これらの情報４００、５００、６００及び７００が、キャッシュメモリ１３０、チャネルアダプタ１１０、ディスクアダプタ１２０、又は物理的記憶装置１５０のいずれかに集中して、又はこれらに分散して格納されてもよい。

【００６４】

図１０は、本実施形態においてデータの複製を開始する手順を説明するフローチャートである。図１０を参照して、正記憶システム１００Ａから副記憶システム１００Ｂに対して、データ複製を開始する手順を説明する。

【００６５】

（１）　グループ作成（ステップ９００）

ユーザは、保守端末あるいはホストコンピュータ１８０を使用して、正記憶システム１００Ａ内のグループ情報６００を参照し、グループ状態が“未使用”のグループ番号、例えば「Ａ」、を取得する。ユーザは、保守端末あるいはホストコンピュータ１８０を使用して、そのグループ番号「Ａ」を指定したグループ作成指示を、正記憶システム１００Ａに入力する。グループ作成指示を受けて、正記憶システム１００Ａは、指定されたグループ番号「Ａ」のグループ状態を“正常”に変更する。

【００６６】

同様に、ユーザは、副記憶システム１００Ｂのグループ情報６００を参照し、グループ状態が“未使用”のグループ番号、例えば「Ｂ」、を取得する。ユーザは、保守端末あるいはホストコンピュータ１８０を使用して、副記憶システム１００Ｂとグループ番号「Ｂ」を指定したグループ作成指示を、正記憶システム１００Ａに入力する。正記憶システム１００Ａは、受信したグループ作成指示を副記憶システム１００Ｂに転送する。副記憶システム１００Ｂは、指定されたグループ番号「Ｂ」のグループ状態を“正常”に変更する。

【００６７】

或いは、ユーザは、副記憶システム１００Ｂの保守端末あるいは副記憶システム１００Ｂに接続されたホストコンピュータ１８０を使用して、グループ番号「Ｂ」を指定したグループ作成指示を、副記憶システム１００Ｂに直接入力してもよい。この場合も、副記憶システム１００Ｂは、指定されたグループ番号「Ｂ」のグループ状態を“正常”に変更す

る。

【００６８】

（２）ペア登録（ステップ９１０）

ユーザは、保守端末あるいはホストコンピュータ１８０を使用して、データ複製元である正論理ボリュームを特定する情報とデータ複製先である副論理ボリュームを特定する情報を指定したペア登録指示を、正記憶システム１００Ａに入力する。正論理ボリュームを特定する情報には、作成されたグループ番号「Ａ」と、そのグループ「Ａ」内の正論理ボリューム番号が含まれる。副論理ボリュームを特定する情報には、副記憶システム１００Ｂの記憶システム番号と、作成されたグループ番号「Ｂ」と、そのグループ「Ｂ」内の副論理ボリューム番号が含まれる。

【００６９】

前記ペア登録指示を受けて、正記憶システム１００Ａは、ペア情報５００からペア情報が“未使用”のペア番号を取得する。そして、正記憶システム１００Ａ、ペア情報５００内の取得したペア番号の行において、ペア状態を“コピー未”に設定し、正記憶システム番号に、正記憶システム１００Ａの記憶システム番号を設定し、正論理ボリューム番号に、指示された正論理ボリューム番号を設定し、副記憶システム番号に、指示された副記憶システム番号を設定し、副論理ボリューム番号に、指示された副論理ボリューム番号を設定し、そして、グループ番号に、指示されたグループ番号「Ａ」を設定する。さらに、正記憶システム１００Ａは、指示されたグループ番号「Ａ」のグループ情報６００のペア集合に、前記取得したペア番号を追加し、正論理ボリューム番号のボリューム状態を“正”に変更する。

【００７０】

正記憶システム１００Ａは、正記憶システム１００Ａの記憶システム番号、ユーザから指定されたグループ番号「Ｂ」、正論理ボリューム番号、および副論理ボリューム番号を、副記憶システム１００Ｂに通知する。副記憶システム１００Ｂは、ペア情報５００から未使用のペア番号を取得する。そして、副記憶システム１００Ｂは、ペア情報５００内の取得したペア番号の行において、ペア状態を“コピー未”に設定し、正記憶システム番号に、記憶システム１００Ａの記憶システム番号を設定し、正論理ボリューム番号に、指示された正論理ボリューム番号を設定し、副記憶システム番号に、副記憶システム１００Ｂの副記憶システム番号を設定し、副論理ボリューム番号に、指示された副論理ボリューム番号を設定し、そして、グループ番号に、指示されたグループ番号「Ｂ」を設定する。さらに、副記憶システム１００Ｂは、指示されたグループ番号「Ｂ」のグループ情報６００のペア集合に、前記取得したペア番号を追加し、副論理ボリューム番号のボリューム状態を“副”に変更する。

【００７１】

以上の動作が、全てのデータ複製対象のペアに対して行われる。

【００７２】

前記の説明では、論理ボリュームのグループへの登録と、論理ボリュームのペアの設定を同時に行う処理を説明したが、それぞれ個別に行ってもよい。

【００７３】

（３）ジャーナル論理ボリューム登録（ステップ９２０）

ユーザは、保守端末あるいはホストコンピュータ１８０を使用して、ジャーナルの保存に使用する論理ボリューム（ジャーナル論理ボリューム）をグループに登録する指示（ジャーナル論理ボリューム登録指示）を、正記憶システム１００Ａに入力する。ジャーナル論理ボリューム登録指示には、グループ番号と論理ボリューム番号が含まれる。

【００７４】

正記憶システム１００Ａは、指示されたグループ番号のグループ情報６００内のジャーナル論理ボリューム番号に、指示された論理ボリューム番号を登録し、そして、当該論理ボリュームのボリューム情報４００内のボリューム状態に、“正常”を設定する。

【００７５】

同様に、ユーザは、保守端末あるいはホストコンピュータ１８０を使用して、副記憶システム１００Ｂのボリューム情報４００を参照し、副記憶システム１００Ｂの記憶システム番号、グループ番号「Ｂ」、ジャーナル論理ボリュームの論理ボリューム番号を指定して、ジャーナル論理ボリューム登録の指示を正記憶システム１００Ａに入力する。正記憶システム１００Ａは、そのジャーナル論理ボリューム登録指示を副記憶システム１００Ｂに転送する。副記憶システム１００Ｂは、指示されたグループ番号「Ｂ」のグループ情報６００内のジャーナル論理ボリューム番号に、指示された論理ボリューム番号を登録し、そして、当該論理ボリュームのボリューム情報４００内のボリューム状態に、“正常”を設定する。

【００７６】

ユーザは、副記憶システム１００Ｂの保守端末あるいは副記憶システム１００Ｂに接続したホストコンピュータ１８０を使用して、グループ番号、ジャーナル論理ボリュームの論理ボリューム番号を指定して、ジャーナル論理ボリューム登録指示を副記憶システム１００Ｂに直接入力してもよい。この場合も、副記憶システム１００Ｂは、指示されたグループ番号「Ｂ」のグループ情報６００内のジャーナル論理ボリューム番号に、指示された論理ボリューム番号を登録し、そして、当該論理ボリュームのボリューム情報４００内のボリューム状態に、“正常”を設定する。

【００７７】

以上の動作が、ジャーナル論理ボリュームとして使用される全ての論理ボリュームに関して行われる。ステップ９１０とステップ９２０の順は逆であってもよい。

【００７８】

（４）　データ複製処理の開始（ステップ９３０）

ユーザは、保守端末あるいはホストコンピュータ１８０を使用して、データ複製処理を開始するグループ番号を指定して、データ複製処理の開始を正記憶システム１００Ａに指示する。正記憶システム１００Ａは、指示されたグループに属する全てのペア情報４００のコピー済みアドレスを“０”に設定する。

【００７９】

正記憶システム１００Ａは、副記憶システム１００Ｂに、ジャーナルリード処理２４０およびリストア処理２５０（図２参照）の開始を指示する。正記憶システム１００Ａは、後述する初期コピー処理を開始する。

【００８０】

（５）初期コピー処理の終了（ステップ９４０）

初期コピー処理が終了すると、正記憶システム１００Ａは、初期コピー処理の終了を副記憶システム１００Ｂに通知する。副記憶システム１００Ｂは、指示されたグループに属する全ての副論理ボリュームのペア状態を“正常”に変更する。

【００８１】

図１１は、上述した図１０のステップ９３０で行われる初期コピー処理の手順を示すフローチャートである。

【００８２】

初期コピー処理では、データ複製対象の正論理ボリュームの全記憶領域に関して、ペア情報５００のコピー済みアドレスを用いながら、その全記憶領域の先頭から順に、単位データサイズ毎にジャーナルが作成される。コピー済みアドレスは、初期値は０であり、そこに、ジャーナルが作成される都度、ジャーナル作成が済んだデータのデータサイズが加算される。論理ボリュームの記憶領域の先頭から、コピー済みアドレスの直前のアドレスまでの範囲については、初期コピー処理にてジャーナルが作成済みである。初期コピー処理を行うことにより、正論理ボリューム内の更新されていないデータを副論理ボリュームに転送することが可能となる。以下の説明では、正記憶システム１００Ａ内のチャネルアダプタ１１０（図１参照）が初期コピー処理を行うように説明されるが、代わって、ディスクアダプタ１２０がそれを行ってもよい。以下、図１１に基づいて初期コピー処理の手順を説明する。

【００８３】
（１）　正記憶システム１００Ａ内のチャネルアダプタ１１０は、処理対象のグループに属するペアの中でペア状態が“コピー未”である正論理ボリューム、例えば「Ａ」を得、そのペアの状態を“コピー中”に変更し、以下の処理を繰り返す（ステップ１０１０、１０２０）。もし、“コピー未”のペア状態にある正論理ボリュームが存在しない場合は、処理を終了する（ステップ１０３０）。
【００８４】
（２）　ステップ１０１０にて取得した正論理ボリューム「Ａ」について、チャネルアダプタ１１０は、単位サイズ（例えば、１ＭＢ）のデータ毎に、ジャーナルを作成する（ステップ１０４０）。ジャーナル作成処理の詳細については後述する
【００８５】
（３）　チャネルアダプタ１１０は、ペア情報５００のコピー済みアドレスに、ジャーナル作成が済んだデータのサイズを加算する（ステップ１０５０）。
【００８６】
（４）　コピー済みアドレスが、正論理ボリューム「Ａ」の容量に達するまで、上記ステップ１０４０及び１０５０の処理を繰り返す（ステップ１０６０）。コピー済みアドレスが、正論理ボリューム「Ａ」の容量と等しくなった場合、正論理ボリューム「Ａ」の全記憶領域に対してジャーナルが作成されたことになるため、チャネルアダプタ１１０は、ペア状態を“正常”に更新し、そして、他の正論理ボリュームについて初期コピー処理を開始する（ステップ１０７０）。
【００８７】
図１１のフローチャートでは、個々の正論理ボリュームが逐次に処理されるが、複数の論理ボリュームが同時に処理されてもよい。
【００８８】
図１２は、命令受信処理２１０（図２参照）におけるデータの流れを説明する図である。図１３は、命令受信処理２１０の手順を示すフローチャートである。図１４は、ジャーナル作成処理のフローチャートである。以下、図１２～図１４を参照して、正記憶システム１００Ａが、ホストコンピュータ１８０からデータ複製対象の論理ボリューム２３０にライト命令を受信した場合の動作について説明する。
【００８９】
まず、図１２及び図１３を参照して、命令受信処理２１０の全体的な動作を説明する。
【００９０】
（１）　正記憶システム１００Ａ内のチャネルアダプタ１１０は（図１３のステップ１２００）、ホストコンピュータからアクセス命令を受信する。アクセス命令は、データリード、データライト、又は後述するジャーナルリード等の命令、命令対象の論理アドレス、及びデータ量等を含んでいる。以下の説明では、アクセス命令内の論理アドレスを論理アドレス「Ａ」、論理ボリューム番号を論理ボリューム「Ａ」、論理ボリューム内位置を論理ボリューム内位置「Ａ」、データ量をデータ量「Ａ」とする。
【００９１】
（２）　チャネルアダプタ１１０は、アクセス命令を調べる（ステップ１２１０、１２１５）。ステップ１２１５の調べで、アクセス命令がジャーナルリード命令の場合は、後述するジャーナルリード受信処理を行う（ステップ１２２０）。アクセス命令がジャーナルリード命令およびデータライト命令以外、例えば、データリード命令の場合は、従来技術と同じようにデータリード処理を行う（ステップ１２３０）。
【００９２】
（３）　ステップ１２１０の調べで、アクセス命令がデータライト命令の場合は、チャネルアダプタ１１０は、論理ボリュームＡのボリューム情報４００を参照し、ボリューム状態を調べる（ステップ１２４０）。ステップ１２４０の調べで、論理ボリュームＡのボリューム状態が、“正常”もしくは“正”以外の場合は、論理ボリュームＡへのアクセスは不可能なため、チャネルアダプタ１１０は、ホストコンピュータ１８０に異常終了を報

告する（ステップ１２４５）。

【００９３】

（４）　ステップ１２４０の調べで、論理ボリュームＡのボリューム状態が、“正常”、“正”のいずれかの場合は、チャネルアダプタ１１０は、キャッシュメモリ１３０を確保し、ホストコンピュータ１８０にデータ受信の準備ができたことを通知する。ホストコンピュータ１８０は、その通知を受け、ライトデータを正記憶システム１００Ａに送信する。チャネルアダプタ１１０は、ライトデータを受信し、これを当該キャッシュメモリ１３０に保存する（ステップ１２５０、図１２の１１００）。

【００９４】

（５）　チャネルアダプタ１１０は、論理ボリューム「Ａ」のボリューム状態を参照し、論理ボリューム「Ａ」がデータ複製対象かどうかを調べる（ステップ１２６０）。ステップ１２６０の調べで、ボリューム状態が、“正”である場合は、論理ボリューム「Ａ」がデータ複製対象であるため、後述するジャーナル作成処理を行う（ステップ１２６５）。

【００９５】

（６）　ステップ１２６０の調べで、ボリューム状態が、“正常”である場合、もしくはステップ１２６５のジャーナル作成処理の終了後、チャネルアダプタ１１０は、ディスクアダプタ１２０にライトデータを記憶装置１５０に書き込むことを命令し（図１２の１１４０）、ホストコンピュータ１８０に終了報告する（ステップ１２７０、１２８０）。その後、当該ディスクアダプタ１２０は、リードライト処理により、物理的記憶装置１５０にライトデータを保存する（図１２の１１１０）。

【００９６】

次に、上述した図１１のステップ１０４０又は図１３のステップ１２６５のジャーナル作成処理について、図１２及び図１４を参照して説明する。

【００９７】

（１）　チャネルアダプタ１１０は、最初のセットのポインタ情報７００を取得する（図１４のステップ１３０５）。チャネルアダプタ１１０は、ジャーナル論理ボリュームのボリューム状態を調べる（ステップ１３１０）。ステップ１３１０の調べで、ジャーナル論理ボリュームのボリューム状態が、“異常”の場合は、ジャーナル論理ボリュームにジャーナルの格納が不可能なため、チャネルアダプタ１１０は、グループ状態を“異常”に変更し、処理を終了する（ステップ１３１５）。この場合、チャネルアダプタ１１０は、ジャーナル論理ボリュームを正常な論理ボリュームに変更する等を行う。

【００９８】

（２）　ステップ１３１０の調べで、ジャーナル論理ボリュームが正常である場合、チャネルアダプタ１１０は、ジャーナル作成処理を継続する。ジャーナル作成処理は、初期コピー処理内の処理であるか、命令受信処理内の処理であるかによって動作が異なる（ステップ１３２０）。ジャーナル作成処理が命令受信処理内の処理の場合は、ステップ１３３０からの動作が行われる。ジャーナル作成処理が初期コピー処理内の場合は、ステップ１３７０からの動作が行われる。

【００９９】

（３）　ジャーナル作成処理が命令受信処理内の処理の場合、チャネルアダプタ１１０は、ライト対象の論理アドレス「Ａ」が、初期コピー処理の処理対象となったかを調べる（ステップ１３３０）。論理ボリューム「Ａ」のペア状態が“コピー未”の場合は、後に初期コピー処理にてジャーナル作成処理が行われるため、チャネルアダプタ１１０は、ジャーナルを作成せずに処理を終了する（ステップ１３３５）。論理ボリューム「Ａ」のペア状態が“コピー中”の場合は、コピー済みアドレスが論理アドレス内位置「Ａ」と等しいか小さいならば、後に初期コピー処理にてジャーナル作成処理が行われるため、チャネルアダプタ１１０は、ジャーナルを作成せずに処理を終了する（ステップ１３３５）。上記以外、つまり、論理ボリューム「Ａ」のペア状態が“コピー中”かつコピー済みアドレスが論理アドレス内位置「Ａ」以上の場合もしくは、論理ボリューム「Ａ」のペア状態が

“正常”の場合は、既に初期コピー処理が完了しているため、チャネルアダプタ１１０は、ジャーナル作成処理を継続する。

【０１００】

（４）　次に、チャネルアダプタ１１０は、ジャーナルがジャーナル論理ボリュームに格納可能であるかを調べる。すなわち、チャネルアダプタ１１０は、ポインタ情報７００を用い、更新情報領域の未使用領域の有無を調べる（ステップ１３４０）。ポインタ情報７００の更新情報最新アドレスと更新情報最古アドレスが等しい場合は、更新情報領域に未使用領域が存在しないため、チャネルアダプタ１１０は、ジャーナル作成失敗として処理を終了する（ステップ１３９０）。

【０１０１】

ステップ１３４０の調べで、更新情報領域に未使用領域が存在する場合は、チャネルアダプタ１１０は、ポインタ情報７００を用い、ライトデータ領域にライトデータが格納できるかを調べる（ステップ１３４５）。ライトデータ最新アドレスとデータ量「Ａ」の和が、ライトデータ最古アドレスと等しいか大きい場合は、ライトデータ領域に格納できないため、チャネルアダプタ１１０は、ジャーナル作成失敗として処理を終了する（ステップ１３９０）。次のセットのポインタ情報７００が存在する限り、チャネルアダプタ１１０は、次セットのポインタ情報７００について、ステップ１３４０～ステップ１３４５の処理を繰り返す。

【０１０２】

（５）　ジャーナルが格納可能である場合、チャネルアダプタ１１０は、更新番号と、更新情報を格納するための論理アドレスと、ライトデータを格納するための論理アドレスを取得し、更新情報をキャッシュメモリ１３０内に作成する。チャネルアダプタ１１０は、対象グループのグループ情報６００から更新番号を取得し、そして、これに１を加算した数値を、グループ情報６００の更新番号に設定する。チャネルアダプタ１１０は、更新情報を格納するための論理アドレスとして、ポインタ情報７００の更新情報最新アドレスを取得し、そして、それに更新情報のサイズを加算した数値を、ポインタ情報７００の更新情報最新アドレスに設定する。チャネルアダプタ１１０は、ライトデータを格納するための論理アドレスとして、ポインタ情報７００のライトデータ最新アドレスを取得し、そして、このライトデータ最新アドレスにデータ量「Ａ」を加算した数値を、ポインタ情報７００のライトデータ最新アドレスに設定する。チャネルアダプタ１１０は、上記取得した数値と、グループ番号と、ライト命令を受信した時刻と、ライト命令内の論理アドレス「Ａ」及びデータ量「Ａ」を、更新情報に設定する（ステップ１３５０、図１２の１１２０）。例えば、図７に例示したグループ情報６００及び図８に例示したポインタ情報７００が存在する場合において、グループ「１」に属する正論理ボリューム「１」の記憶領域の先頭からアドレス数「８００」だけ進んだ位置にデータサイズ「１００」のデータをライトせよというライト命令を受信した場合、チャネルアダプタ１１０は、図１５に例示するような更新情報を作成することになる。そして、図７に例示したグループ情報６００内の更新番号は「５」に変更される。また、ポインタ情報６００内の更新情報最新アドレスは「６００」（更新情報のサイズを「１００」とした場合）に変更され、ライトデータ最新アドレスは「２３００」に変更される。正記憶システムに接続された複数の副記憶システムにそれぞれ対応する複数セットのポインタ情報７００が存在する場合、チャネルアダプタ１１０は、それら複数セットのポインタ情報７００の全てを上記のようにして更新する。

【０１０３】

（６）　チャネルアダプタ１１０は、ディスクアダプタ１２０に、ジャーナルの更新情報とライトデータを記憶装置１５０に書き込むことを命令し、正常終了する（ステップ１３６０、図１２の１１３０、１１４０、１１５０）。

【０１０４】

（７）　ジャーナル作成処理が、初期コピー処理内の処理の場合は、ステップ１３７０からの動作が行われる。チャネルアダプタ１１０は、ジャーナルが作成可能であるかを調

べる。すなわち、チャネルアダプタ１１０は、ポインタ情報７００を用い、更新情報領域の未使用領域の有無を調べる（ステップ１３７０）。ポインタ情報７００の更新情報最新アドレスと更新情報最古アドレスが等しい場合は、更新情報領域に未使用領域が存在しないため、チャネルアダプタ１１０は、ジャーナル作成失敗として処理を終了する（ステップ１３９０）。ここで、本実施形態における初期コピー処理においては、ジャーナルのライトデータは、正論理ボリュームからリードされ、ライトデータ領域は使用されないため、ライトデータ領域の未使用領域の確認は不要である。次のセットのポインタ情報７００が存在する限り、チャネルアダプタ１１０は、次セットのポインタ情報７００についてステップ１３７０の処理を繰り返す。

【０１０５】

（８）　ステップ１３７０の調べで、ジャーナルが作成可能である場合、チャネルアダプタ１１０は、更新情報に設定すべき以下のような数値を取得し、更新情報をキャッシュメモリ１３０内に作成する。すなわち、チャネルアダプタ１１０は、対象グループのグループ情報６００から更新番号を取得し、そして、これに１を足した数値をグループ情報６００の更新番号に設定する。チャネルアダプタ１１０は、更新情報を格納するための論理アドレスとして、ポインタ情報７００内の更新情報最新アドレスの位置を取得し、そして、これに更新情報のサイズを足した数値をポインタ情報７００内の更新情報最新アドレスに設定する。

【０１０６】

チャネルアダプタ１１０は、上記取得した数値と、グループ番号と、本処理の開始時刻と、初期コピー処理対象の論理アドレスと、初期コピーの１回の処理データ量と、初期コピー処理におけるライトデータを格納したジャーナル論理ボリュームの論理アドレスとを、更新情報に設定する（ステップ１３８０、図１２の１１２０）。

【０１０７】

（９）　チャネルアダプタ１１０は、ディスクアダプタ１２０に、更新情報を物理的記憶装置１５０に書き込むことを命令し、正常終了する（ステップ１３８５、図１２の１１４０、１１５０）。

【０１０８】

上記説明では、更新情報はキャッシュメモリ１３０内に格納されるが、変形として、更新情報が共有メモリ１４０又はその他の記憶場所に格納されてもよい。

【０１０９】

ディスクアダプタ１２０によるライトデータの物理的記憶装置１５０への書き込みは、非同期で行うことができる（つまり、ステップ１３６０およびステップ１３８５の直後でなくともよい）。ただし、ホストコンピュータ１８０が、論理アドレス「Ａ」に対するライト命令を再び発行した場合には、ジャーナルのライトデータが上書きされることになる。そのため、ホストコンピュータ１８０からライトデータを受信する前に、ジャーナルのライトデータは、更新情報の示すジャーナル論理ボリュームの論理アドレスに対応する物理記憶装置１５０内の記憶場所に、書き込まれる必要がある。もしくは、ジャーナルのライトデータは、別のキャッシュメモリに退避されて、後に更新情報のジャーナル論理ボリュームの論理アドレスに対応する物理的記憶装置１５０内の記憶場所に書き込まれてもよい。

【０１１０】

前述したジャーナル作成処理では、ジャーナルは適当なタイミングでキャッシュメモリから物理的記憶装置１５０に保存される。変形例として、ジャーナル用に予め一定量のキャッシュメモリ１３０を用意しておき、当該キャッシュメモリが全て使用された段階で、ジャーナルが物理的記憶装置１５０に保存されるようにしてもよい。ジャーナル用のキャッシュメモリ量は、例えば、保守端末から指定できるようにしてよい。

【０１１１】

リードライト処理２２０（図２、図１２参照）は、ディスクアダプタ１２０が、チャネルアダプタ１１０もしくはディスクアダプタ１２０から命令を受け、実施する処理である

。リードライト処理２２０は、指定されたキャッシュメモリ１３０のデータを、指定された論理アドレスに対応する物理的記憶装置１５０内の記憶領域に書き込むライト処理、及び、指定された論理アドレスに対応する記憶装置１５０内の記憶領域のデータを、指定されたキャッシュメモリ１３０に読み出すリード処理等に分類される。

【０１１２】

図１６は、ジャーナルリード命令を受信した正記憶システム１００Ａのチャネルアダプタ１１０の動作（ジャーナルリード受信処理）におけるデータ流れを説明する図です。図１７は、ジャーナルリード受信処理の手順を示すフローチャートである。以下、これらの図面を用いて、正記憶システム１００Ａが、副記憶システム１００Ｂからジャーナルリード命令を受信した場合の動作について説明する。

【０１１３】

（１）　図１６及び図１７に示すように、正記憶システム１００Ａ内のチャネルアダプタ１１０は、副記憶システム１００Ｂからアクセス命令（ジャーナルリード命令）を受信する。このアクセス命令は、ジャーナルリード命令であることを示す識別子、命令対象のグループ番号、及びリトライ指示の有無などを゛含んでいる（ステップ１２２０、図１６の１４１０）。以下の説明では、そのアクセス命令（ジャーナルリード命令）内のグループ番号をグループ番号「Ａ」とする。

【０１１４】

（２）　チャネルアダプタ１１０は、グループ番号「Ａ」のグループ状態が“正常”であるかを調べる（ステップ１５１０）。ステップ１５１０の調べで、グループ状態が“正常”以外、例えば、“障害”の場合は、チャネルアダプタ１１０は、副記憶システム１００Ｂにグループ状態を通知し、処理を終了する。副記憶システム１００Ｂは、受信したグループ状態に応じて処理を行う。例えば、副記憶システム１００Ｂは、グループ状態が“障害”の場合は、ジャーナルリード処理を終了する（ステップ１５１５）。

【０１１５】

（３）　ステップ１５１０の調べで、グループ番号「Ａ」のグループ状態が“正常”の場合、チャネルアダプタ１１０は、ジャーナル論理ボリュームの状態を調べる（ステップ１５２０）。ステップ１５２０の調べで、ジャーナル論理ボリュームのボリューム状態が“正常”でない場合、例えば、“障害”の場合は、チャネルアダプタ１１０は、グループ状態を“障害”に変更し、副記憶システム１００Ｂにグループ状態を通知し、処理を終了する。副記憶システム１００Ｂは、受信したグループ状態に応じて処理を行う。例えば、副記憶システム１００Ｂは、グループ状態が“障害”の場合は、ジャーナルリード処理を終了する（ステップ１５２５）。

【０１１６】

（４）　ステップ１５２０の調べで、ジャーナル論理ボリュームのボリューム状態が“正常”の場合は、チャネルアダプタ１１０は、ジャーナルリード命令がリトライ指示かを調べる（ステップ１５３０）。

【０１１７】

（５）　ステップ１５３０の調べで、ジャーナルリード命令がリトライ指示の場合、チャネルアダプタ１１０は、前回送信したジャーナルを再度、副記憶システム１００Ｂに送信する。チャネルアダプタ１１０は、キャッシュメモリ１３０を確保し、ディスクアダプタ１２０に、ポインタ情報７００のリトライ開始アドレスが指す記憶場所から、更新情報のサイズ分のデータ(更新情報）をキャッシュメモリに読み込むことを命令する（図１６の１４２０）。

【０１１８】

ディスクアダプタ１２０のリードライト処理２２０は、物理的記憶装置１５０から更新情報を読み込み、これをキャッシュメモリ１３０に保存し、更新情報のリード終了をチャネルアダプタ１１０に通知する（図１６の１４３０）。

【０１１９】

チャネルアダプタ１１０は、更新情報のリード終了の通知を受け、更新情報から、ライ

トデータの論理アドレスおよびライトデータのサイズを取得し、キャッシュメモリ１３０を確保し、ディスクアダプタ１２０にライトデータをキャッシュメモリ１３０に読み込むことを命令する（ステップ１５４０、図１６の１４４０）。

【０１２０】

ディスクアダプタ１２０のリードライト処理２２０は、物理的記憶装置１５０からライトデータを読み込み、これをキャッシュメモリ１３０に保存し、ライトデータのリード終了をチャネルアダプタ１１０に通知する（図１６の１４５０）。

【０１２１】

チャネルアダプタ１１０は、ライトデータのリード終了の通知を受け、更新情報とライトデータを副記憶システム１００Ｂに送信し、ジャーナルを保持しているキャッシュメモリ１３０を開放し、処理を終了する（ステップ１５４５、図１６の１４６０）。

【０１２２】

（６）　ステップ１５３０の調べで、リトライ指示でない場合、チャネルアダプタ１１０は、送信していないジャーナルが存在するかを調べ、存在すれば、そのジャーナルを副記憶システム１００Ｂに送信する。チャネルアダプタ１１０は、ポインタ情報７００のリード開始アドレスと更新情報最新アドレスを比較する（ステップ１５５０）。

【０１２３】

リード開始アドレスが更新情報最新アドレスと等しい場合は、全てのジャーナルを副記憶システム１００Ｂに送信済みであるため、チャネルアダプタ１１０は、副記憶システム１００Ｂに“ジャーナル無“を送信し（ステップ１５６０）、そして、前回のジャーナルリード命令の処理時に副記憶システム１００Ｂに送信したジャーナルの記憶領域を開放する（ステップ１５９０）。

【０１２４】

ステップ１５９０のジャーナルの記憶領域の開放処理では、その記憶領域に対して積極的にこれを開放する動作を行う必要はなく、単に最古アドレスのポインタを更新するだけでよい。すなわち、ポインタ情報７００の更新情報最古アドレスに、リトライ開始アドレスが設定される。更新情報最古アドレスがライトデータ領域先頭アドレスとなった場合は、更新情報最古アドレスは“０”とされる。ポインタ情報７００のライトデータ最古アドレスは、その現在値に、前回のリードジャーナル命令に応じて送信されたライトデータのサイズを加算した数値に変更される。ライトデータ最古アドレスが、ジャーナル論理ボリュームの容量以上の論理アドレスとなった場合は、ライトデータ領域先頭アドレスが減じられるよう補正される。

【０１２５】

（７）　ステップ１５５０の調べで、未送信のジャーナルが存在する場合、チャネルアダプタ１１０は、キャッシュメモリ１３０を確保し、ディスクアダプタにポインタ情報７００のリード開始アドレスから、更新情報のサイズ分のデータ（更新情報）をキャッシュメモリに読み込むことを命令する（図１６の１４２０）。

【０１２６】

ディスクアダプタ１２０のリードライト処理２２０は、物理的記憶装置１５０から更新情報を読み込み、これをキャッシュメモリ１３０に保存し、更新情報のリード終了をチャネルアダプタ１１０に通知する（図１６の１４３０）。

【０１２７】

チャネルアダプタ１１０は、更新情報のリード終了の通知を受け、更新情報から、ライトデータの論理アドレスおよびライトデータのサイズを取得し、キャッシュメモリ１３０を確保し、ディスクアダプタ１２０にライトデータをキャッシュメモリに読み込むことを命令する（ステップ１５７０、図１６の１４４０）。

【０１２８】

ディスクアダプタ１２０のリードライト処理２２０は、物理的記憶装置１５０からライトデータを読み込み、これをキャッシュメモリ１３０に保存し、ライトデータのリード終了をチャネルアダプタに通知する（図１６の１４５０）。

【０１２９】
チャネルアダプタ１１０は、ライトデータのリード終了の通知を受け、更新情報とライトデータを副記憶システム１００Ｂに送信し（ステップ１５８０）、ジャーナルを保持しているキャッシュメモリ１３０を開放する（図１６の１４６０）。そして、ポインタ情報７００のリトライ開始アドレスにリード開始アドレスを設定し、そして、リード開始アドレスを、その現在値に、送信されたジャーナルの更新情報サイズを加算した数値に変更する。
【０１３０】
（８）　チャネルアダプタ１１０は、前回のジャーナルリード命令の処理時に副記憶システム１００Ｂに送信されたジャーナルの記憶領域を開放する（ステップ１５９０）。
【０１３１】
上述したジャーナルリード受信処理では、正記憶システム１００Ａは、ジャーナルを一つずつ逐次に副記憶システム１００Ｂに送信する。変形例として、複数のジャーナルを同時に副記憶システム１００Ｂに送信するようにしてもよい。１つのジャーナルリード命令で送信されるジャーナル数は、副記憶システム１００Ｂによりジャーナルリード命令内で指定されてもよいし、或いは、グループ登録の時等に、ユーザにより正記憶システム１００Ａ又は副記憶システム１００Ｂに指定されてもよい。さらに、正記憶システム１００Ａと副記憶システム１００Ｂの接続パス２００の転送能力又は負荷等に応じて動的に、１つのジャーナルリード命令で送信されるジャーナル数が変更されてもよい。また、ジャーナル数でなく、ジャーナルのライトデータのサイズに応じて、ジャーナルの転送量が指定されてもよい。
【０１３２】
前述したジャーナルリード受信処理では、ジャーナルが物理的記憶装置１５０からキャッシュメモリ１３０に読み込まれる。しかし、この動作は、ジャーナルが既にキャッシュメモリ１３０に存在する場合は不要である。
【０１３３】
前述したジャーナルリード受信処理内のジャーナルの記憶領域の開放処理は、次のジャーナルリード命令の処理時に開放される。変形例として、副記憶システム１００Ｂにジャーナルを送信した直後に、そのジャーナルの記憶領域が開放されてもよい。また、副記憶システム１００Ｂが、ジャーナルリード命令内で開放してよい更新番号を設定し、正記憶システム１００Ａは、その設定により指定された更新番号のジャーナルの記憶領域を開放するようにしてもよい。
【０１３４】
図１８は、ジャーナルリード（ＪＮＬＲＤ）処理２４０におけるデータ流れを説明する図である。図１９は、ジャーナルリード処理２４０の手順を示すフローチャートである。図２０は、ジャーナルリード処理２４０の中で行われるジャーナル格納処理の手順を示すフローチャートである。以下、これらの図面を用いて、副記憶システム１００Ｂのチャネルアダプタ１１０が、正記憶システム１００Ａからジャーナルを読み出し、ジャーナル論理ボリュームに格納する動作について説明する。
【０１３５】
（１）　副記憶システム１００Ｂ内のチャネルアダプタ１１０は、ジャーナルを格納するキャッシュメモリ１３０を確保し、ジャーナルリード命令であることを示す識別子、命令対象の正記憶システム１００Ａのグループ番号、及びリトライ指示の有無を含むアクセス命令（ジャーナルリード命令）を、正記憶システム１００Ａに送信する（ステップ１７００、図１８の１６１０）。以下、アクセス命令内のグループ番号をグループ番号「Ａ」とする。
【０１３６】
（２）　チャネルアダプタ１１０は、正記憶システム１００Ａの応答およびジャーナルを受信する（図１８の１６２０）。チャネルアダプタ１１０は、正記憶システム１００Ａからの応答を調べ、その応答が、“ジャーナル無”の場合は、正記憶システム１００Ａに

は、指定したグループのジャーナルが存在しないため、一定時間後、正記憶システム１００Ａにリードジャーナル命令を送信する（ステップ１７２０、１７２５）。

【０１３７】

（４）　正記憶システム１００Ａの応答が、“グループ状態は障害”もしくは“グループ状態は未使用”の場合は、チャネルアダプタ１１０は、副記憶システム１００Ｂのグループ状態を受信した状態に変更し、ジャーナルリード処理を終了する（ステップ１７３０、１７３５）。

【０１３８】

（５）　正記憶システム１００Ａの応答が、上記以外、つまり、正常終了の場合は、チャネルアダプタ１１０は、ジャーナル論理ボリュームのボリューム状態を調べる（ステップ１７４０）。ジャーナル論理ボリュームのボリューム状態が“異常”の場合は、ジャーナル論理ボリュームにジャーナルの格納が不可能なため、チャネルアダプタ１１０は、グループ状態を“異常”に変更し、処理を終了する（ステップ１７４５）。この場合、チャネルアダプタ１１０は、ジャーナル論理ボリュームを正常な論理ボリュームに変更する等を行い、グループの状態を正常に戻す。

【０１３９】

（６）　ステップ１７４０の調べで、ジャーナル論理ボリュームのボリューム状態が“正常”の場合は、チャネルアダプタ１１０は、後述するジャーナル格納処理１８００を行う。ジャーナル格納処理１８００が正常に終了した場合は、チャネルアダプタ１１０は、直ちに次のジャーナルリード命令を送信するか、もしくは、一定時間経過後、次のジャーナルリード命令を送信する（ステップ１７６０）。次のジャーナル命令を送信するタイミングは、一定の時間間隔で定期的に送信するものでもよく、その時間間隔は、受信したジャーナルの個数、接続線２００の通信量、副記憶システム１００Ｂが保持しているジャーナルの記憶容量、又は副記憶システム１００Ｂの負荷等に応じて制御されてもよい。さらに、正記憶システム１００Ａが保持しているジャーナルの記憶容量又は正記憶システム１００Ａのポインタ情報を、副記憶システム１００Ｂが読み出し、その数値に基づいて、上記時間間隔が制御されてもよい。上記情報の転送は、専用のコマンドで行われてもよいし、ジャーナルリード命令の応答に含まれていてよい。その後の処理は、ステップ１７１０以降と同じである。

【０１４０】

（７）　ステップ１８００のジャーナル格納処理が正常に終了しない場合は、ジャーナル論理ボリュームの未使用領域が足りないため、チャネルアダプタ１１０は、受信したジャーナルを破棄し、一定時間後にリトライ指示のジャーナルリード命令を送信する（ステップ１７５５）。もしくは、チャネルアダプタ１１０は、ジャーナルをキャッシュメモリに保持しておき、一定時間後に、再度ジャーナル格納処理を行う。これは、後述するリストア処理２５０が行われることにより、一定時間後には、ジャーナル論理ボリュームに未使用領域が増える可能性があるためである。この方式の場合は、ジャーナルリード命令にリトライ指示の有無は不要である。

【０１４１】

次に、上述したジャーナル格納処理１８００について、図２０を参照して説明する。

【０１４２】

（１）　副記憶システムチャネルアダプタ１１０は、最初のセットのポインタ情報７００を取得する（ステップ１８０５）。副記憶システムチャネルアダプタ１１０は、ジャーナルがジャーナル論理ボリュームに格納可能であるかを調べる。すなわち、チャネルアダプタ１１０は、ポインタ情報７００を用い、更新情報領域に未使用領域の有無を調べる（ステップ１８１０）。ポインタ情報７００の更新情報最新アドレスと更新情報最古アドレスが等しい場合は、更新情報領域に未使用領域が存在しないため、チャネルアダプタ１１０は、ジャーナル作成失敗として処理を終了する（ステップ１８２０）。

【０１４３】

（２）　ステップ１８１０の調べで、更新情報領域に未使用領域が存在する場合は、チ

ャネルアダプタ１１０は、ポインタ情報７００を用い、ライトデータ領域にライトデータが格納できるかを調べる（ステップ１８３０）。ライトデータ最新アドレスと受信したジャーナルのライトデータのデータ量の和が、ライトデータ最古アドレスと等しいもしくは大きい場合は、ライトデータ領域にライトデータを格納できないため、チャネルアダプタ１１０は、ジャーナル作成失敗として処理を終了する（ステップ１８２０）。次のセットのポインタ情報７００が存在する限り、チャネルアダプタ１１０は、次セットのポインタ情報７００についてステップ１８１０～ステップ１８３０を繰り返す。

【０１４４】

（３）　ジャーナルが格納可能である場合、チャネルアダプタ１１０は、受信した更新情報のグループ番号とジャーナル論理ボリュームの論理アドレスを変更する。グループ番号は、副記憶システム１００Ｂのグループ番号に変更され、ジャーナル論理ボリュームの論理アドレスはポインタ情報７００のライトデータ最新アドレスに変更される。チャネルアダプタ１１０は、ポインタ情報７００の更新情報最新アドレスを、これに更新情報最新アドレスに更新情報のサイズを足した数値に変更する。チャネルアダプタ１１０は、ポインタ情報７００のライトデータ最新アドレスを、このライトデータ最新アドレスにライトデータのサイズを足した数値に変更する（ステップ１８４０）。複数の副記憶システムにそれぞれ対応する複数セットのポインタ情報が存在する場合、チャネルアダプタ１１０は、それら複数セットのポインタ情報７００の全てを上記のように更新する。

【０１４５】

（４）　チャネルアダプタ１１０は、ディスクアダプタ１２０に、更新情報とライトデータを記憶装置１５０に書き込むことを命令し、ジャーナル作成成功として処理を終了する（ステップ１８５０、図１８の１６３０）。その後、ディスクアダプタ１２０は、リードライト処理２２０により、物理的記憶装置１５０に更新情報とライトデータを書き込み、キャッシュメモリ１３０を開放する（図１８の１６４０）。

【０１４６】

上述したジャーナル格納処理では、ジャーナルが適当なタイミングでキャッシュメモリ１３０から物理的記憶装置１５０に保存される。変形例として、ジャーナル用に予め一定量のキャッシュメモリ１３０を用意しておき、当該キャッシュメモリが全て使用された段階で、物理的記憶装置１５０にジャーナルを保存してもよい。ジャーナル用のキャッシュメモリ量は、例えば、保守端末から指定できるようにしてよい。

【０１４７】

図２１は、リストア処理２５０におけるデータ流れを説明する図である。図２２は、リストア処理２５０の手順を示すフローチャートである。以下、これらの図面を用いて、副記憶システム１００Ｂのチャネルアダプタ１１０が、ジャーナルを利用し、データの更新を行う動作について説明する。変形例として、リストア処理２５０が、副記憶システム１００Ｂのディスクアダプタ１２０により行われてもよい。

【０１４８】

（１）　副記憶システム１００Ｂチャネルアダプタ１１０は、グループ番号Ｂのグループ状態が“正常”であるかを調べる（ステップ２０１０）。ステップ２０１０の調べで、グループ状態が“正常”以外、例えば、“障害”の場合は、チャネルアダプタ１１０は、リストア処理を終了する（ステップ２０１５）。

【０１４９】

（２）　ステップ２０１０の調べで、グループ状態が“正常”の場合は、チャネルアダプタ１１０は、ジャーナル論理ボリュームのボリューム状態を調べる（ステップ２０２０）。ステップ２０２０の調べで、ジャーナル論理ボリュームのボリューム状態が、“異常”の場合は、アクセス不可能なため、チャネルアダプタ１１０は、グループ状態を“異常”に変更し、処理を終了する（ステップ２０２５）。

【０１５０】

（３）　ステップ２０２０の調べで、ジャーナル論理ボリュームのボリューム状態が、“正常”の場合は、チャネルアダプタ１１０は、リストア対象のジャーナルが存在するか

を調べる。すなわち、チャネルアダプタ１１０は、ポインタ情報７００の更新情報最古アドレスと更新情報最新アドレスを取得する。更新情報最古アドレスと更新情報最新アドレスが等しい場合、ジャーナルは存在しないため、チャネルアダプタ１１０は、リストア処理は一旦終了し、一定時間後、リストア処理を再開する（ステップ２０３０）。

【０１５１】

（４）　ステップ２０３０の調べで、リストア対象のジャーナルが存在する場合、チャネルアダプタ１１０は、最古（最小）の更新番号を持つジャーナルに対して次の処理を行う。最古（最小）の更新番号を持つジャーナルの更新情報は、ポインタ情報７００の更新情報最古アドレスから保存されている。チャネルアダプタ１１０は、キャッシュメモリ１３０を確保し、ディスクアダプタ１２０に、更新情報最古アドレスから、更新情報のサイズ分のデータ(更新情報）をキャッシュメモリ１３０に読み込むことを命令する（図２１の１９１０）。

【０１５２】

ディスクアダプタ１２０のリードライト処理２２０は、物理的記憶装置１５０から更新情報を読み込み、キャッシュメモリ１３０に保存し、更新情報のリード終了をチャネルアダプタ１１０に通知する（図２１の１９２０）。

【０１５３】

チャネルアダプタ１１０は、更新情報のリード終了の通知を受け、更新情報から、ライトデータの論理アドレスおよびライトデータのサイズを取得し、キャッシュメモリ１３０を確保し、ディスクアダプタ１２０にライトデータをキャッシュメモリに読み込むことを命令する（図２１の１９３０）。

【０１５４】

ディスクアダプタ１２０のリードライト処理２２０は、物理的記憶装置１５０からライトデータを読み込み、キャッシュメモリ１３０に保存し、ライトデータのリード終了をチャネルアダプタ１１０に通知する（ステップ２０４０、図２１の１９４０）。

【０１５５】

（５）　チャネルアダプタ１１０は、更新情報から更新する副論理ボリュームの論理アドレスを求め、ディスクアダプタ１２０に副論理ボリュームにライトデータを書き込むことを命令する（ステップ２０５０、図２１の１９５０）。ディスクアダプタ１２０のリードライト処理２２０は、副論理ボリュームの論理アドレスに対応する物理的記憶装置１５０の記憶場所にデータを書き込み、キャッシュメモリ１３０を開放し、データのライト完了をチャネルアダプタ１１０に通知する（図２１の１９６０）。

【０１５６】

（６）　チャネルアダプタ１１０は、データのライト完了の通知を受け、ジャーナルの記憶領域を開放する。ジャーナルの記憶領域の開放処理では、ポインタ情報７００の更新情報最古アドレスが、これに更新情報のサイズを足した数値に変更される。更新情報最古アドレスが、ライトデータ領域先頭アドレスとなった場合は、更新情報最古アドレスは“０”とされる。ポインタ情報７００のライトデータ最古アドレスは、これにライトデータのサイズを足した数値に変更される。ライトデータ最古アドレスが、ジャーナル論理ボリュームの容量以上の論理アドレスとなった場合は、ライトデータ領域先頭アドレスが減じされるよう補正される。その後、チャネルアダプタ１１０は、次のリストア処理を開始する（ステップ２０６０）。

【０１５７】

前述したリストア処理２５０では、物理的記憶装置１５０からキャッシュメモリ１３０にジャーナルが読み込まれる。しかし、この動作は、ジャーナルが既にキャッシュメモリ１３０に存在する場合は不要である。

【０１５８】

前述したジャーナルリード受信処理とジャーナルリード処理２４０では、正記憶システム１００Ａが、送信すべきジャーナルをポインタ情報７００により決めている。変形例として、副記憶システム１００Ｂが、送信すべきジャーナルを決めてもよい。この場合、副

記憶システム１００Ｂは、例えば、ジャーナルリード命令に、更新番号を追加することができる。この場合、正記憶システム１００Ａの共有メモリ１４０内に、更新番号から更新情報を格納した論理アドレスを求めるためのテーブルもしくは検索方法が設けられ、それにより、正記憶システム１００Ａが、ジャーナルリード受信処理にて、副記憶システム１００Ｂにより指定された更新番号から、更新情報の論理アドレスを求めるようにすることができる。

【０１５９】

前述したジャーナルリード受信処理とジャーナルリード処理２４０では、ジャーナルリード命令という専用のアクセス命令が用いられる。変形例として、通常のリード命令を用いるようにしてもよい。その場合、例えば、正記憶システム１００Ａのグループ情報６００とポインタ情報７００が予め副記憶システム１００Ｂに転送され、そして、副記憶システム１００Ｂが、グループ情報６００とポインタ情報７００に基づいて、正記憶システム１００Ａ内のジャーナル論理ボリュームのデータ（つまり、ジャーナル）をリードするためのリード命令を生成するようにすることができる。

【０１６０】

前述したジャーナルリード受信処理では、更新番号の順に、正記憶システム１００Ａから副記憶システム１００Ｂにジャーナルが順次に送信される。変形例として、更新番号の順とは異なる順序でジャーナルが送信されるようにしてもよい。或いは、正記憶システム１００Ａから副記憶システム１００Ｂに、複数のジャーナルを並列的に送信されてもよい。この場合、副記憶システム１００Ｂに、更新番号から更新情報を格納した論理アドレスを求めるテーブルもしくは検索方法が設けられ、それにより、副記憶システム１００Ｂでのリストア処理２５０で、更新番号順にジャーナルが処理されるようにすることができる。

【０１６１】

本実施形態では、正記憶システム１００Ａがジャーナルを取得し、副記憶システム１００Ｂが、正記憶システム１００Ａからジャーナルをリードして、どれに基づいてデータの複製を行う。これにより、正記憶システム１００Ａに接続されたホストコンピュータ１８０は、データの複製に関する負荷を負わない。さらに、正記憶システム１００Ａと副記憶システム１００Ｂ間でジャーナルが転送されるので、正記憶システム１００Ａとホストコンピュータ１８０間の通信線がデータ複製のためには使用されない。

【０１６２】

図２３は、本発明に従うデータ処理システムの第２の実施形態の論理的な構成を示す図である。

【０１６３】

図２３に示すように、この実施形態は、正記憶システム１００Ａと副記憶システム１００Ｂの他に、第３の記憶システム１００Ｃを有する。これらの記憶システム１００Ａ、１００Ｂ、１００Ｃの物理的な構成は、いずれも、図１を参照して既に説明したそれと基本的に同一でよい。ホストコンピュータ１８０と第３の記憶システム１００Ｃが接続パス１９０により接続され、第３の記憶システム１００Ｃと正記憶システム１００Ａが接続パス２００により接続され、そして、正記憶システム１００Ａと副記憶システム１００Ｂが接続パス２００により接続される。第３の記憶システム１００Ｃは、正記憶システム１００Ａ内の正論理ボリューム（「ＤＡＴＡ１」、「ＤＡＴＡ２」等）２３０内のデータの元のデータをそれぞれ保持したオリジナル論理ボリューム（「ＯＲＧ１」、「ＯＲＧ２」等）２３０を有する。

【０１６４】

第３の記憶システム１００Ｃは、ホストコンピュータ１８０からのデータライト命令により、要求されたオリジナル論理ボリューム（例えば「ＯＲＧ１」）２３０内のデータ（オリジナルデータ）を更新する。この時、第３の記憶システム１００Ｃは、オリジナル論理ボリューム（例えば「ＯＲＧ１」）２３０内のオリジナルデータを更新するだけでなく、その更新対象のオリジナルデータに対応する正論理ボリューム（「ＤＡＴＡ１」）２３

０内のデータを更新するためのデータライト命令を、正記憶システム１００Ａに送る（２３１０）。

【０１６５】

正記憶システム１００Ａは、第１の実施形態で説明した通り、上記データライト命令を受けて、要求された正論理ボリューム（例えば「ＤＡＴＡ１」）２３０のデータを更新し、そして、前述した命令受信処理２１０およびリードライト処理２２０によって、そのデータ更新のジャーナルを、ジャーナル論理ボリューム（「ＪＮＬ１」）２３０に保存する（２３１０）。

【０１６６】

副記憶システム１００Ｂは、前述したジャーナルリード処理２４０によって、正記憶システム１００Ａからジャーナルをリードし、リードライト処理２２０によって、ジャーナル論理ボリューム（「ＪＮＬ２」）２３０にジャーナルを保存する（２３２０）。

【０１６７】

正記憶システム１００Ａは、副記憶システム１００Ｂからジャーナルをリードする命令を受信すると、命令受信処理２１０およびリードライト処理２２０によって、ジャーナル論理ボリューム（「ＪＮＬ１」）２３０からジャーナルを読み出し、副記憶システム１００Ｂに送信する（２３２０）。

【０１６８】

副記憶システム１００Ｂは、前述したリストア処理２５０およびリードライト処理２２０によって、更新番号に従い、ジャーナル論理ボリューム（「ＪＮＬ２」）からジャーナルを読み出し、正論理ボリューム（「ＤＡＴＡ１」）２３０の複製である副論理ボリューム（ＣＯＰＹ１）のデータを更新する（２３３０）。このように、更新番号の順にデータを更新することにより、論理ボリューム間のデータの整合性を保つことが可能となる。

【０１６９】

図２３に示したデータ処理システムでは、正記憶システム１００Ａがジャーナルを取得し、これをジャーナル専用の記憶領域に格納する。さらに、副記憶システム１００Ｂは、正記憶システムから受信したジャーナルを、ジャーナル専用の記憶領域に格納する。ジャーナル専用の記憶領域は、データ複製対象の記憶領域より少なくすることが可能であり、より少ない記憶容量で、副記憶システムに正記憶システムのデータの複製が可能となる。

【０１７０】

図２４は、本発明に従うデータ処理システムの第３の実施形態の論理的な構成を示す図である。

【０１７１】

図２４に示すように、この実施形態は、正記憶システム１００Ａと副記憶システム１００Ｂの他に、第３の記憶システム１００Ｃを有する。これらの記憶システム１００Ａ、１００Ｂ、１００Ｃの物理的な構成は、いずれも、図１を参照して既に説明したそれと基本的に同一でよい。ホストコンピュータ１８０と第３の記憶システム１００Ｃが接続パス１９０により接続され、第３の記憶システム１００Ｃと正記憶システム１００Ａが接続パス２００により接続され、そして、正記憶システム１００Ａと副記憶システム１００Ｂが接続パス２００により接続される。

【０１７２】

正記憶システム１００Ａは、第３の記憶システム１００Ｃに対し、正論理ボリューム（「ＤＡＴＡ１」、「ＤＡＴＡ２」等）があるように見せるが、その正論理ボリューム（「ＤＡＴＡ１」、「ＤＡＴＡ２」等）に対しては、実際の物理的な記憶領域、つまり物理的記憶装置１５０を割り当てない。例えば、ボリューム情報４００内の各正論理ボリュームの物理アドレスに、物理的記憶装置１５０を割り当てていないことを示す所定の数値が設定される。従って、正記憶システム１００Ａ内の正論理ボリューム（「ＤＡＴＡ１」、「ＤＡＴＡ２」等）は、仮想的なものである。正記憶システム１００Ａは、これら仮想的な正論理ボリューム（「ＤＡＴＡ１」、「ＤＡＴＡ２」等）のデータ更新のジャーナルを保持するためのジャーナル論理ボリューム（「ＪＮＬ１」）２３０を有し、これには、実際

の物理的な記憶領域が割り当てられる。第３の記憶システム１００Ｃは、正記憶システム１００Ａ内の仮想的な正論理ボリューム（「ＤＡＴＡ１」、「ＤＡＴＡ２」）内のデータに相当する実際のデータを保持したオリジナル論理ボリューム（例えば「ＯＲＧ１」、「ＯＲＧ２」等）２３０を有し、これには実際の物理的記憶領域が割り当てられている。

【０１７３】

第３の記憶システム１００Ｃは、ホストコンピュータ１８０からのデータライト命令により、要求されたオリジナル論理ボリューム（例えば「ＯＲＧ１」）のデータ（オリジナルデータ）を更新する。この時、第３の記憶システム１００Ｃは、そのオリジナルデータを更新するだけでなく、更新されるオリジナルデータに対応する仮想的な正論理ボリューム（例えば「ＤＡＴＡ１」）内のデータを更新するためのデータライト命令を、正記憶システム１００Ａに送る（２４１０）。

【０１７４】

正記憶システム１００Ａは、第３の記憶システム１００Ｃから仮想的な正論理ボリューム（例えば「ＤＡＴＡ１」）内のデータのライト命令を受信すると、図１３に示した命令受信処理２１０のステップ１２７０の処理（ディスクアダプタへのデータライト命令の発行）は行わずに、そのデータ更新のジャーナルをジャーナル論理ボリューム（「ＪＮＬ１」）２３０に保存するだけである(２４１０)。

【０１７５】

副記憶システム１００Ｂは、前述したジャーナルリード処理２４０によって、正記憶システム１００Ａからジャーナルをリードし、リードライト処理２２０によって、ジャーナル論理ボリューム（「ＪＮＬ２」）２３０にジャーナルを保存する(２４２０)。

【０１７６】

正記憶システム１００Ａは、副記憶システム１００Ｂからジャーナルリード命令を受信すると、命令受信処理２１０およびリードライト処理２２０によって、ジャーナル論理ボリューム（「ＪＮＬ１」）２３０からジャーナルを読み出し、記憶システム１００Ｂに送信する(２４２０)。

【０１７７】

副記憶システム１００Ｂは、前述したリストア処理２５０およびリードライト処理２２０によって、更新番号に従い、ジャーナル論理ボリューム（「ＪＮＬ２」）２３０からジャーナルを読み出し、オリジナル論理ボリューム（例えば「ＯＲＧ１」）２３０の複製である副論理ボリューム（例えば「ＣＯＰＹ１」）２３０のデータを更新する(２４３０)。このように、更新番号の順にデータを更新することにより、論理ボリューム間のデータの整合性を保つことが可能となる。

【０１７８】

図２４に示したデータ処理システムでは、第３の記憶システム１００Ｃもしくは、第３の記憶システム１００Ｃに接続されたホストコンピュータ１８０に障害が生じた場合、副記憶システム１００Ｂの論理ボリューム（例えば「ＣＯＰＹ１」）２３０に対し、正記憶システム１００Ａ内のジャーナル（「ＪＮＬ１」）２３０を反映することにより、副記憶システム１００Ｂに接続されたホストコンピュータ（図２４には図示せず）により、最新データの参照、更新が可能となる。さらに、正記憶システム１００Ａにオリジナルデータの複製を保持せず、ジャーナルのみを格納することで、データ複製に必要な記憶容量が少なくすることが可能となる。

【０１７９】

図２５は、本発明に従うデータ処理システムの第４の実施形態の論理的な構成を示す図である。

【０１８０】

図２５に示すように、この実施形態は、正記憶システム１００Ａと複数（例えば２台）の副記憶システム１００Ｂ、１００Ｃを有する。これらの記憶システム１００Ａ、１００Ｂ、１００Ｃの物理的な構成は、いずれも、図１を参照して既に説明したそれと基本的に同一でよい。ホストコンピュータ１８０と正記憶システム１００Ａが接続パス１９０によ

り接続され、正記憶システム１００Ａと第１の副記憶システム１００Ｂが接続パス２００により接続され、そして、正記憶システム１００Ａと第２の副記憶システム１００Ｃが接続パス２００により接続される。正記憶システム１００Ａでは、所定の複数の正論理ボリューム（例えば「ＤＡＴＡ１」、「ＤＡＴＡ２」）２３０とジャーナル論理ボリューム（例えば「ＪＮＬ１」）２３０が一つのグループ「グループ１」を構成する。第１の副記憶システム１００Ｂでは、上述した「グループ１」に属する複数の正論理ボリューム（「ＤＡＴＡ１」、「ＤＡＴＡ２」）２３０のそれぞれの複製である複数の副論理ボリューム（例えば「ＣＯＰＹ１」、「ＣＯＰＹ２」）２３０とジャーナル論理ボリューム（例えば「ＪＮＬ２」）２３０が同じグループ「グループ１」を構成する。同様に、第２の副記憶システム１００Ｃでも、「グループ１」の正論理ボリューム（「ＤＡＴＡ１」、「ＤＡＴＡ２」）２３０のそれぞれの複製である複数の副論理ボリューム（例えば「ＣＯＰＹ３」、「ＣＯＰＹ４」）２３０とジャーナル論理ボリューム（例えば「ＪＮＬ３」）２３０が同じグループ「グループ１」を構成する。

【０１８１】

１台の正記憶システム１００Ａに対して複数台の副記憶システム１００Ｂ、１００Ｃが存在するため、図２６に例示するように、複数台の副記憶システム１００Ｂ、１００Ｃにそれぞれ対応した複数のポインタ情報７００Ｂ、７００Ｃが、正記憶システム１００Ａに保持される。第１の副記憶システム１００Ｂには、第１の副記憶システム１００Ｂ用のポインタ情報７００Ｂが保持され、第２の副記憶システム７００Ｃにも、第２の副記憶システム７００Ｃ用のポインタ情報７００Ｃが保持される。これらのポインタ情報７００Ｂ、７００Ｃの各々の構成とその意味するところは、図８及び図９を参照して既に説明したものと同様である。異なる副記憶システム１００Ｂ、１００Ｃは、それぞれ独自にスケジューリングしたタイミングで、ジャーナルリード処理２４０及びリストア処理２５０を行う。そのため、図２７に例示するように、異なる副記憶システム１００Ｂ、１００Ｃ用のポインタ情報７００Ｂ、７００Ｃが指し示すアドレスは必ずしも同じではない。

【０１８２】

再び図２５を参照して、この実施形態の動作を以下に説明する。

【０１８３】

正記憶システム１００Ａは、ホストコンピュータ１８０から或る正論理ボリューム（例えば「ＤＡＴＡ１」）２３０のデータのライト命令を受信すると、前述した命令受信処理２１０およびリードライト処理２２０によって、正論理ボリューム（「ＤＡＴＡ１」）２３０内の要求されたデータを更新し、そして、ジャーナル論理ボリューム（「ＪＮＬ１」）にそのデータ更新のジャーナルを保存する（２５１０）。

【０１８４】

第１の副記憶システム１００Ｂは、前述したジャーナルリード処理２４０によって、正記憶システム１００Ａからジャーナルをリードし、リードライト処理２２０によって、ジャーナル論理ボリューム（「ＪＮＬ２」）にそのジャーナルを保存する（２５２０）。ここで、第１の副記憶システム１００Ｂがジャーナルをリードするタイミング（ジャーナルリード命令を正記憶システム１００Ａに送るタイミング）は、第１の副記憶システム１００Ｂのチャネルアダプタ１１０によって独自にスケジューリングされる。このジャーナルリードのタイミングじゃ図１９のステップ１７６０に関して既に説明したように、例えば、前のジャーナルのジャーナル格納処理１８００の正常終了後に直ちにでもよいし、前のジャーナルのジャーナル格納処理１８００の正常終了後一定時間経過後でもよいし、或いは、一定の時間間隔で定期的にでもよい。定期的にジャーナルリード命令を送る場合には、その時間間隔は、受信したジャーナルの個数、接続線２００の通信量、第１の副記憶システム１００Ｂが保持しているジャーナルの記憶容量、又は第１の副記憶システム１００Ｂの負荷等に応じて制御されてもよい。さらに、正記憶システム１００Ａが保持しているジャーナルの記憶容量又は正記憶システム１００Ａのポインタ情報を、第１の副記憶システム１００Ｂが読み出し、その数値に基づいて、上記時間間隔が制御されてもよい。上記情報の転送は、専用のコマンドで行われてもよいし、ジャーナルリード命令の応答に含まれ

ていてよい。いずれにしても、ジャーナルをリードするタイミングが他の副記憶システム１００Ｃと同期する必要はない。

【０１８５】

正記憶システム１００Ａは、副記憶システム１００Ｂからジャーナルをリードする命令を受信すると、命令受信処理２１０およびリードライト処理２２０によって、ジャーナル論理ボリューム（「ＪＮＬ１」）からジャーナルを読み出し、第１の副記憶システム１００Ｂに送信する(２５２０)。

【０１８６】

第１の副記憶システム１００Ｂは、前述したリストア処理２５０およびリードライト処理２２０によって、更新番号に従い、ジャーナル論理ボリューム（「ＪＮＬ２」）からジャーナルを読み出し、正論理ボリューム（「ＤＡＴＡ１」）の複製である副論理ボリューム（「ＣＯＰＹ１」）のデータを更新する(２９０)。このように、更新番号の順にデータを更新することにより、論理ボリューム間のデータの整合性を保つことが可能となる。

【０１８７】

第２の副記憶システム１００Ｃは、前述したジャーナルリード処理２４０によって、正記憶システム１００Ａからジャーナルをリードし、リードライト処理２２０によって、ジャーナル論理ボリューム（「ＪＮＬ３」）にジャーナルを保存する(２５２０)。ここで、第２の副記憶システム１００Ｃがジャーナルをリードするタイミング（ジャーナルリード命令を正記憶システム１００Ａに送るタイミング）は、第２の副記憶システム１００Ｃのチャネルアダプタ１１０によって独自にスケジューリングされる。このジャーナルリードのタイミングは、図１９のステップ１７６０に関して既に説明したように、例えば前のジャーナルのジャーナル格納処理１８００の正常終了後に直ちにでもよいし、前のジャーナルのジャーナル格納処理１８００の正常終了後一定時間経過後でもよいし、或いは、一定の時間間隔で定期的にでもよい。定期的にジャーナルリード命令を送る場合には、その時間間隔は、受信したジャーナルの個数、接続線２００の通信量、第２の副記憶システム１００Ｃが保持しているジャーナルの記憶容量、又は第２の副記憶システム１００Ｃの負荷等に応じて制御されてもよい。さらに、正記憶システム１００Ａが保持しているジャーナルの記憶容量又は正記憶システム１００Ａのポインタ情報を、第２の副記憶システム１００Ｃが読み出し、その数値に基づいて、上記時間間隔が制御されてもよい。上記情報の転送は、専用のコマンドで行われてもよいし、ジャーナルリード命令の応答に含まれていてよい。いずれにしても、ジャーナルをリードするタイミングは、他の副記憶システム１００Ｂと同期する必要はない。

【０１８８】

正記憶システム１００Ａは、第２の副記憶システム１００Ｃからジャーナルをリードする命令を受信すると、命令受信処理２１０およびリードライト処理２２０によって、ジャーナル論理ボリューム（「ＪＮＬ１」）からジャーナルを読み出し、第２の記憶システム１００Ｃに送信する(２５２０)。

【０１８９】

第２の副記憶システム１００Ｃは、前述したリストア処理２５０およびリードライト処理２２０によって、更新番号に従い、ジャーナル論理ボリューム（「ＪＮＬ３」）からジャーナルを読み出し、正論理ボリューム（「ＤＡＴＡ１」）の複製である副論理ボリューム（「ＣＯＰＹ３」）のデータを更新する(２９０)。このように、更新番号の順にデータを更新することにより、論理ボリューム間のデータの整合性を保つことが可能となる。

【０１９０】

上述したように、異なる副記憶システム１００Ｂ、１００Ｃが、それぞれ独自にスケジューリングしたタイミングでジャーナルリード処理２４０及びリストア処理２５０を行う。副記憶システム１００Ｂ、１００Ｃは、それぞれ、リストア処理(又はジャーナルリード処理）が終わると、リストア処理（又はジャーナルリード処理）の終了した更新番号を含むリストア処理(又はジャーナルリード処理）終了の通知を、正記憶システム１００Ａに送る。正記憶システム１００Ａは、副記憶システム１００Ｂ、１００Ｃからのリストア

処理(又はジャーナルリード処理）終了の通知に基づいて、どの副記憶システムシステム１００Ｂ、１００Ｃがどの更新番号のリストア処理(又はジャーナルリード処理）を終了したかを示す情報を管理する。そして、正記憶システム１００Ａは、その情報に基づいて、全ての副記憶システム１００Ｂ、１００Ｃにてリストア処理(又はジャーナルリード処理）が終了した更新番号のジャーナルについて、ジャーナル論理ボリューム（「ＪＮＬ１」）２３０内の当該ジャーナルの記憶領域を開放する。副記憶システム１００Ｂ、１００Ｃのいずれか一つでも未だリストア処理(又はジャーナルリード処理）が終了してない更新番号のジャーナルについては、正記憶システム１００Ａは、そのジャーナルをジャーナル論理ボリューム（「ＪＮＬ１」）２３０内に維持し、そのジャーナルの記憶領域を開放しない。

【０１９１】

図２５に例示したデータ処理システムによれば、複数台のうちの一部の副記憶システムで障害が発生しても、他の正常な副記憶システムにより正論理ボリュームの複製が維持されるので、安全性が高い。

【０１９２】

図２５に例示したデータ処理システムでは、一台の正記憶システム１００Ａに対して２台の副記憶システム１００Ｂ、１００Ｃが存在する。変形例として、３台以上の副記憶システム１００Ｂ、１００Ｃが一台の正記憶システム１００Ａに対して設けられてもよい。

【０１９３】

図２５に例示したデータ処理システムでは、複数台の副記憶システム１００Ｂ、１００Ｃが並列的に一台の正記憶システム１００Ａからジャーナルをリードする。変形例として、各副記憶システム１００Ｂ、１００Ｃが、正記憶システム１００Ａからジャーナルをリードする機能の他に、他の副記憶システムからジャーナルをリードする機能も兼ね備えて、正記憶システム１００Ａと他の副記憶システムのどれからジャーナルをリードするかを選択できるようにすることもできる。例えば、正記憶システム１００Ａの負荷が小さいときは、全ての副記憶システム１００Ｂ、１００Ｃが、正記憶システム１００Ａからジャーナルをリードするが、他方、正記憶システム１００Ａの負荷が大きいときには、第1の副記憶システム１００Ｂが正記憶システム１００Ａからジャーナルをリードし、その後に、第２の副記憶システム１００Ｂが、第1の副記憶システム１００Ｂからそのジャーナルをリードするというような制御を行うことができる。

【０１９４】

以上、本発明の幾つかの実施形態を説明した。これらの実施形態によれば、記憶システムの上位の計算機に影響を与えずに、或いは、記憶システムと計算機との間の通信にも影響を与えずに、複数の記憶システム間でデータ転送又はデータの複製をすることが可能である。

【０１９５】

さらに、或る実施形態によれば、複数の記憶システム内に保持するデータ格納領域を少なくすることができる。また、或る実施形態によれば、複数の記憶システムの上位の計算機の業務に大きい影響を与えることのなしに、高速かつ効率的に複数の記憶システム間でデータ転送又はデータの複製をすることができる。

【０１９６】

本発明は、上述した実施形態に限定されるものでなく、その要旨を逸脱しない範囲で種々変更可能であることはいうまでもない。

【図面の簡単な説明】

【０１９７】

【図１】本発明の第１の実施形態のデータ処理システムの物理的構成を示すブロック図である。

【図２】第１実施形態のデータ処理システムの論理的構成を示すブロック図である。

【図３】ジャーナルの更新情報とライトデータの関係を説明する図である。

【図４】ジャーナルに含まれる更新情報の具体例を示す図である。

【図５】第１実施形態のボリューム情報の例を説明する図である。
【図６】第１実施形態のペア情報の例を説明する図である。
【図７】第１実施形態のグループ情報の例を説明する図である。
【図８】第１実施形態のポインタ情報の例を説明する図である。
【図９】ポインタ情報内の項目とジャーナル論理ボリュームとの関係を説明する図である。
【図１０】第１実施形態のデータの複製を開始する手順を説明するフローチャートである。
【図１１】第１実施形態の初期コピー処理を説明するフローチャートである。
【図１２】第１実施形態の命令受信処理におけるデータ流れを説明する図である。
【図１３】第１実施形態の命令受信処理の手順を示すフローチャートである。
【図１４】第１実施形態のジャーナル作成処理の手順を示すフローチャートである。
【図１５】第１実施形態のジャーナル作成処理時の更新情報の例を説明する図である。
【図１６】第１実施形態のジャーナルリード受信処理におけるデータ流れを説明する図である。
【図１７】第１実施形態のジャーナルリード受信処理の手順を示すフローチャートである。
【図１８】第１実施形態のジャーナルリード命令処理におけるデータ流れを説明する図である。
【図１９】第１実施形態のジャーナルリード命令処理の手順を示すフローチャートである。
【図２０】第１実施形態のジャーナル格納処理の手順を示すフローチャートである。
【図２１】第１実施形態のリストア処理におけるデータ流れを説明する図である。
【図２２】第１実施形態のリストア処理の手順を示すフローチャートである。
【図２３】本発明の第２の実施形態の論理的な構成を示す図である。
【図２４】本発明の第３の実施形態の論理的な構成を示す図である。
【図２５】本発明の第４の実施形態の論理的な構成を示す図である。
【図２６】第４実施形態のポインタ情報の例を説明する図である。
【図２７】第４実施形態のポインタ情報とジャーナル論理ボリュームとの関係を説明する図である。

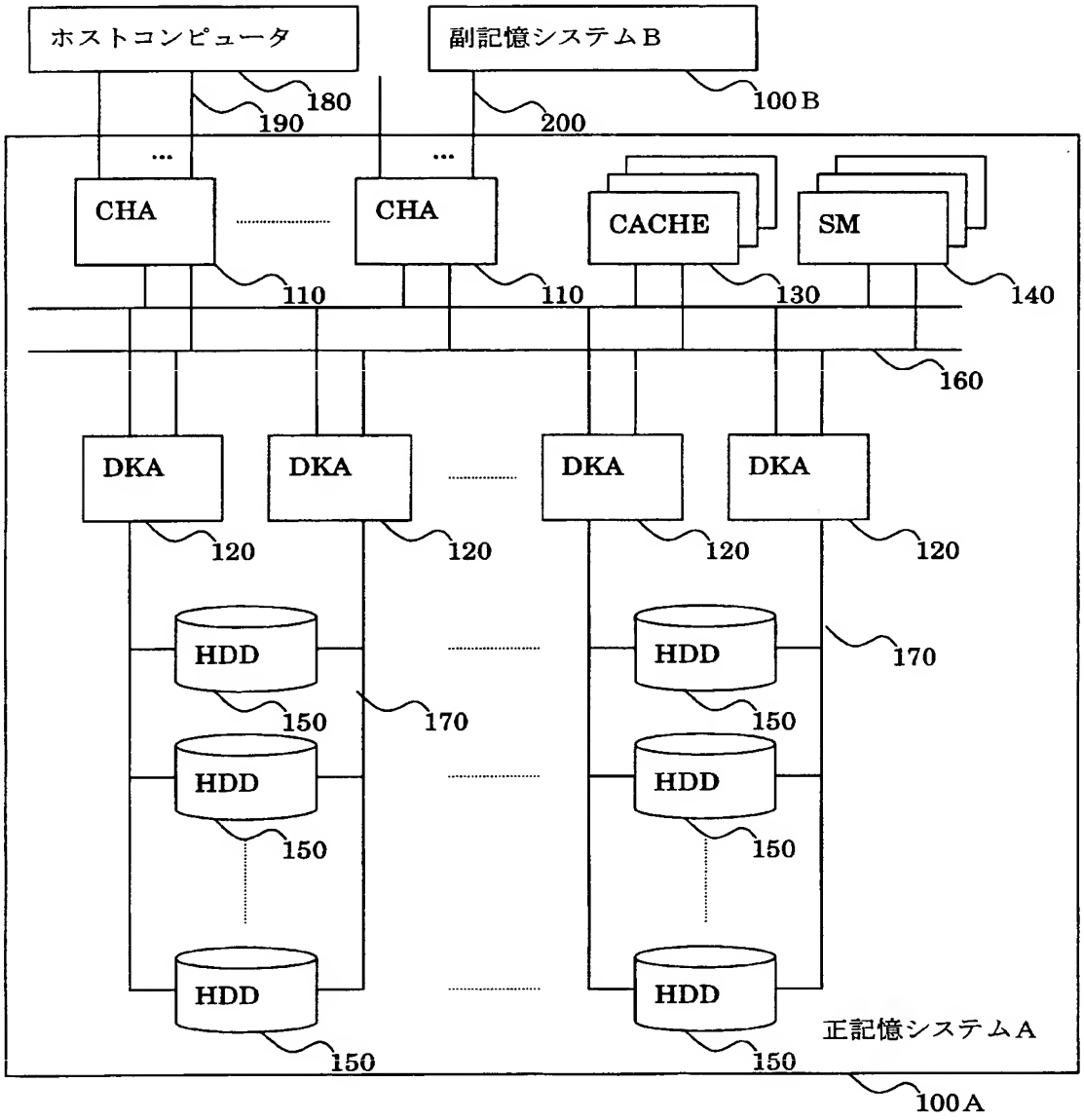
【符号の説明】

【０１９８】

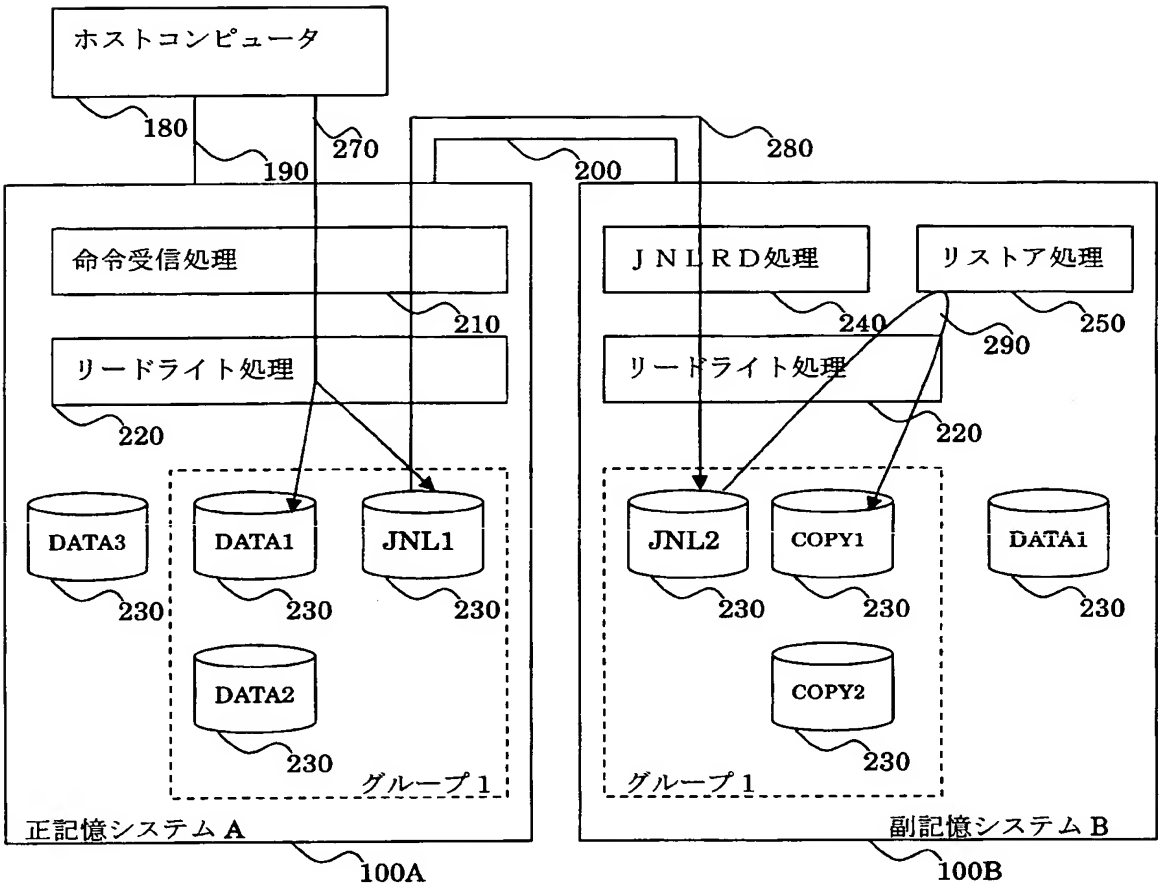
１００　記憶システム
１１０　チャネルアダプタ
１２０　ディスクアダプタ
１３０　キャッシュメモリ
１４０　管理メモリ
１５０　記憶装置
１６０　コモンパス
１７０　ディスクアダプタと記憶装置間の接続線
１８０　ホストコンピュータ

【書類名】図面
【図１】

図１


ホストコンピュータ
副記憶システムＢ
180
190
200
100Ｂ
CHA
CHA
CACHE
SM
110
110
130
140
160
DKA
DKA
DKA
DKA
120
120
120
120
HDD
150
170
正記憶システムＡ
100Ａ

【図２】


ホストコンピュータ
180
190
270
200
280
命令受信処理
210
リードライト処理
220
DATA3
DATA1
JNL1
230
DATA2
グループ１
正記憶システムA
100A
ＪＮＬＲＤ処理
リストア処理
240
290
250
リードライト処理
220
JNL2
COPY1
DATA1
230
COPY2
グループ１
副記憶システムB
100B

【図３】

図３

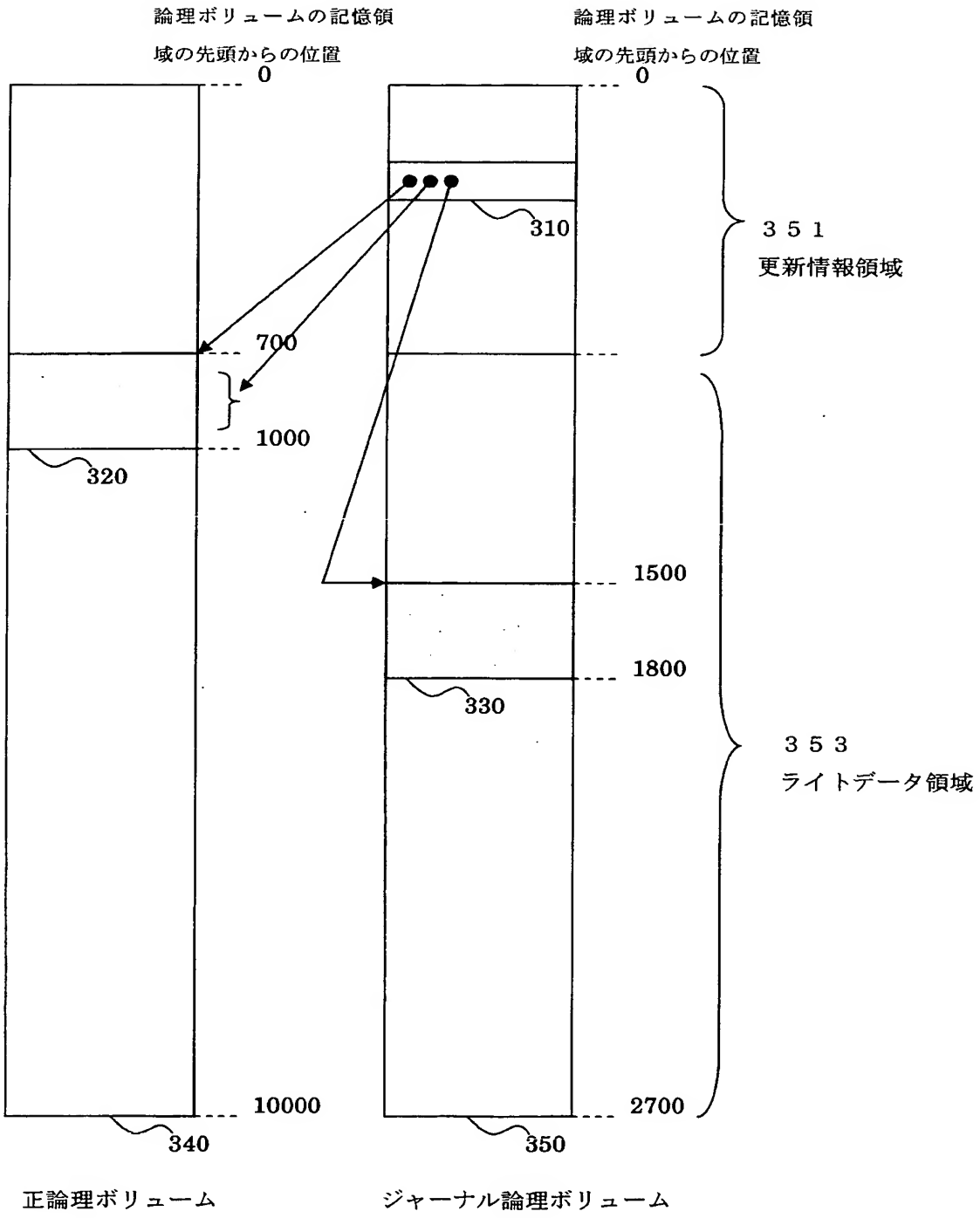

論理ボリュームの記憶領域の先頭からの位置
0
310
３５１
更新情報領域
700
1000
320
1500
1800
330
３５３
ライトデータ領域
10000
340
2700
350
正論理ボリューム
ジャーナル論理ボリューム

図４

設定項目	設定値例
ライト命令を受信した時刻	1999/3/17 22:20:10
グループ番号	1
更新番号	4
ライト命令の論理アドレス	論理ボリューム番号：1 論理ボリュームの記憶領域の先頭からの位置：700
ライトデータのデータサイズ	300
ライトデータを格納したジャーナル論理ボリュームの論理アドレス	論理ボリューム番号：4 論理ボリュームの記憶領域の先頭からの位置：1500

310 更新情報

【図５】

図５

論理ボリューム番号	ボリューム状態	フォーマット形式	容量(GB)	ペア番号	物理アドレス	
					記憶装置番号	先頭から位置
1	正	OPEN3	3	1	1	0
2	正	OPEN6	6	2	1	3
3	未使用	OPEN6	6	0	1	9
4	正常	OPEN9	9	0	2	0
5	正常	OPEN3	3	0	2	9
6	未使用	OPEN6	6	0	2	12

400 ボリューム情報

【図６】

図６

ペア番号	ペア状態	正記憶システム番号	正論理ボリューム番号	副記憶システム番号	副論理ボリューム番号	グループ番号	コピー済みアドレス
1	正常	1	1	2	1	1	0
2	正常	1	2	2	3	1	0
3	未使用	0	0	0	0	0	0
4	未使用	0	0	0	0	0	0
5	未使用	0	0	0	0	0	0

500　ペア情報

【図７】

図７

グループ番号	グループ状態	ペア集合	ジャーナル論理ボリューム番号	更新番号
1	正常	1,2	4	4
2	未使用	0	0	0

600　グループ情報

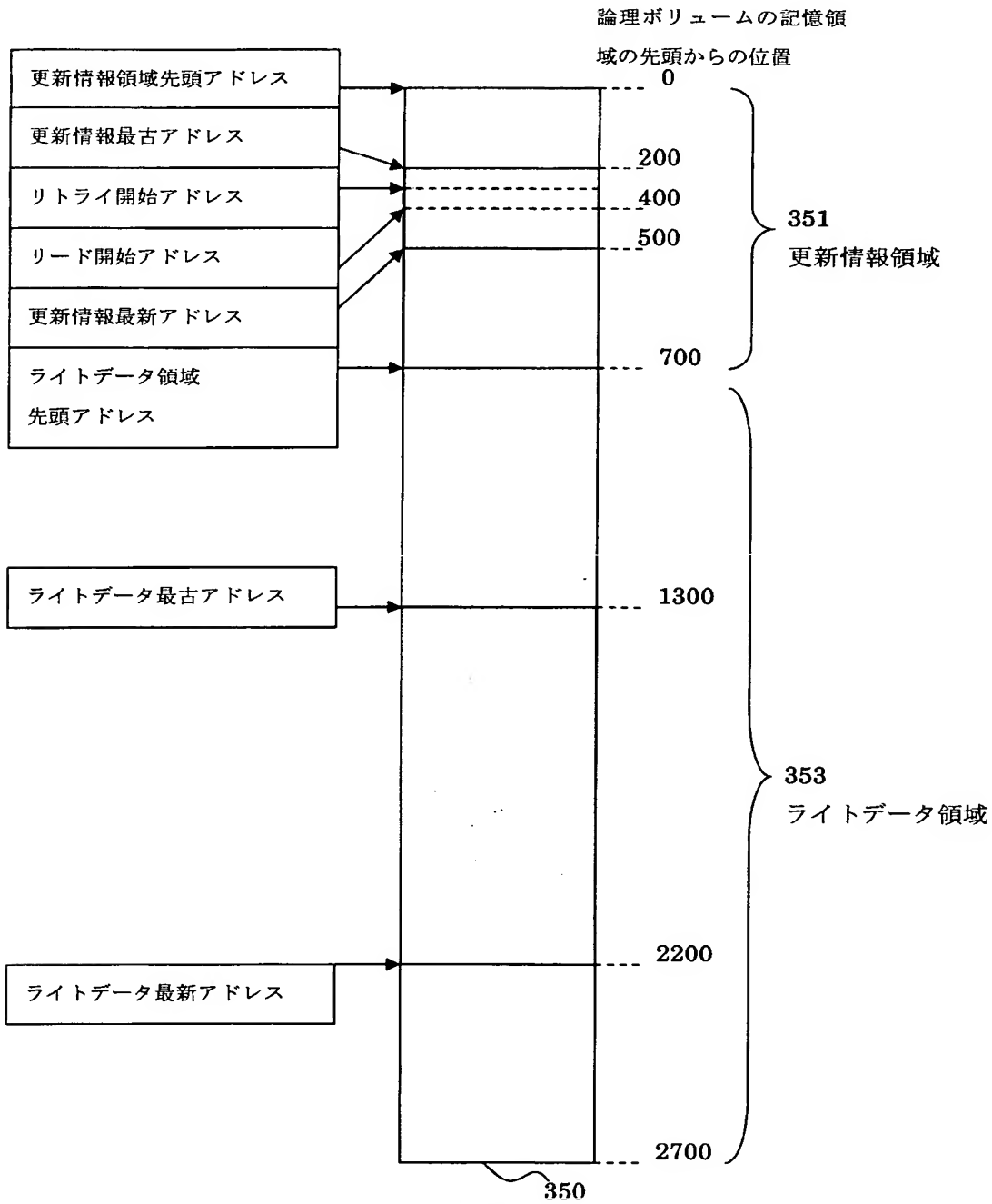
図８

	副記憶システム番号	論理アドレス	
		論理ボリューム番号	論理ボリュームの記憶領域の先頭からの位置
更新情報領域先頭アドレス	2	4	0
ライトデータ領域先頭アドレス	2	4	700
更新情報最新アドレス	2	4	500
更新情報最古アドレス	2	4	200
ライトデータ最新アドレス	2	4	2200
ライトデータ最古アドレス	2	4	1300
リード開始アドレス	2	4	400
リトライ開始アドレス	2	4	300

700　ポインタ情報

【図９】

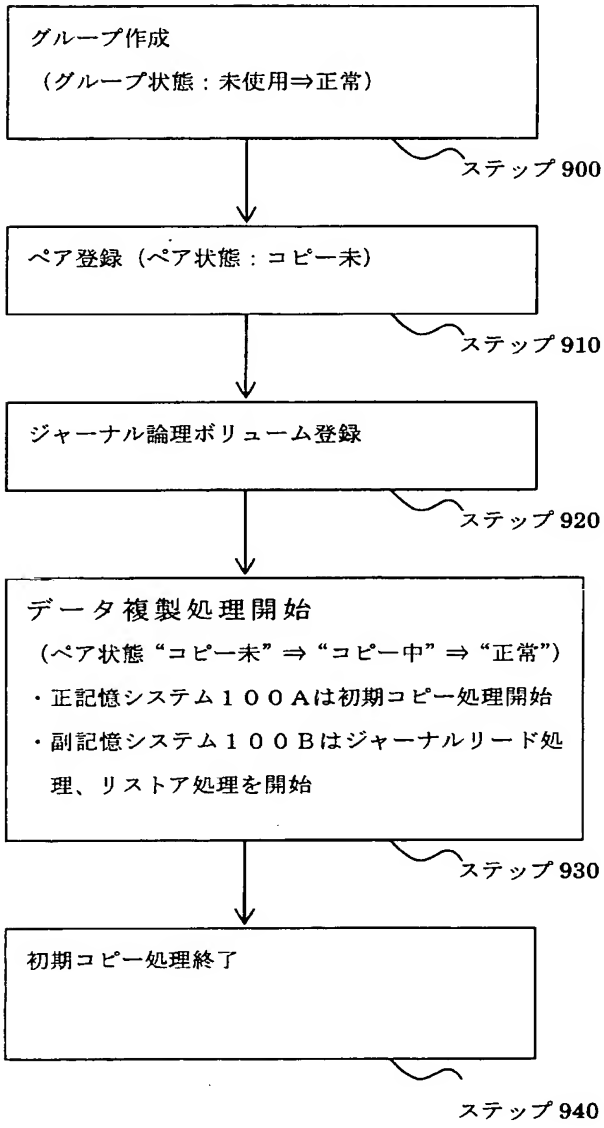
図９



ジャーナル論理ボリューム

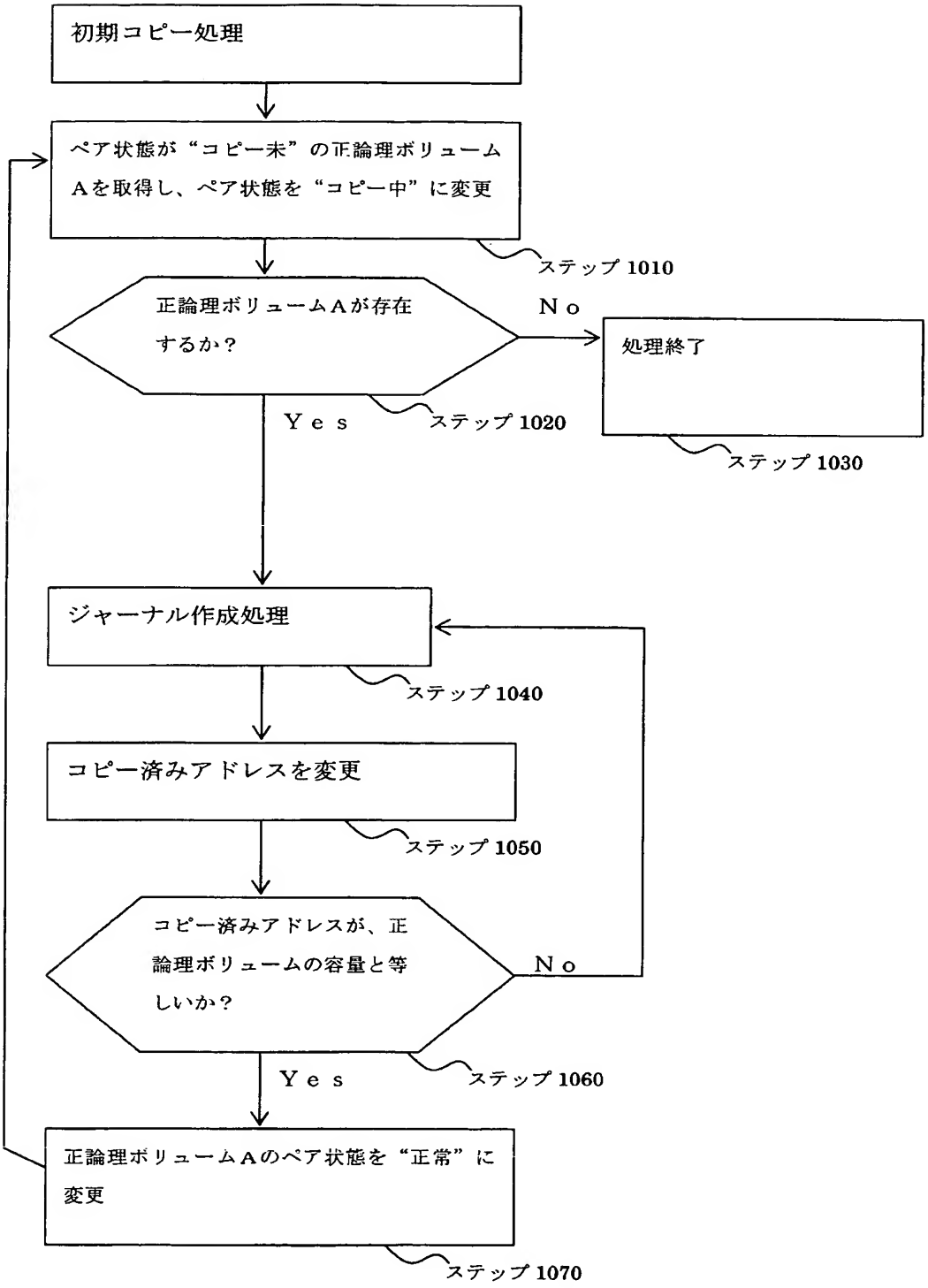
【図１０】

図１０


グループ作成
（グループ状態：未使用⇒正常）
ステップ900
ペア登録（ペア状態：コピー未）
ステップ910
ジャーナル論理ボリューム登録
ステップ920
データ複製処理開始
（ペア状態“コピー未”⇒“コピー中”⇒“正常”）
・正記憶システム１００Ａは初期コピー処理開始
・副記憶システム１００Ｂはジャーナルリード処理、リストア処理を開始
ステップ930
初期コピー処理終了
ステップ940

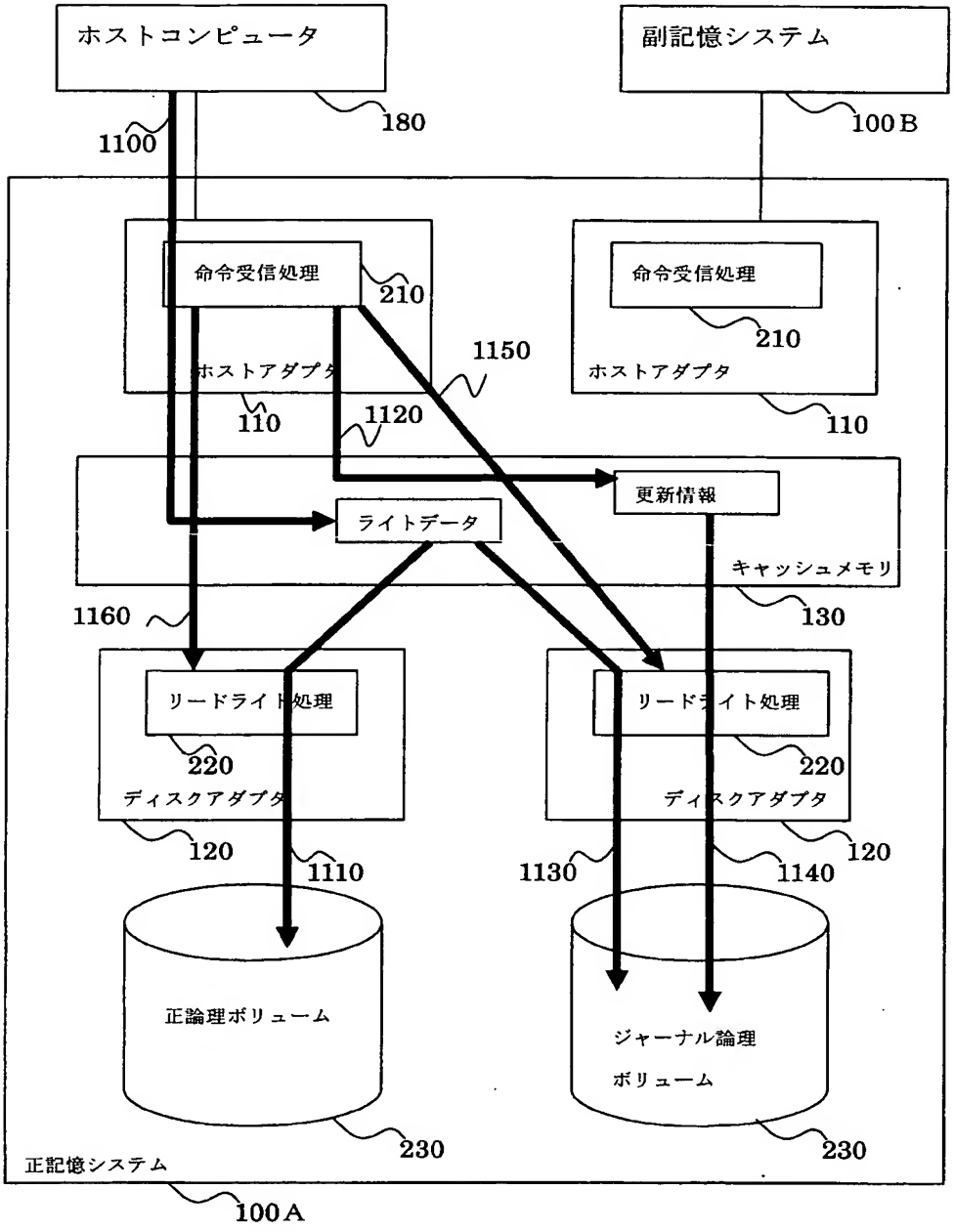
【図１１】

図１１


初期コピー処理
ペア状態が“コピー未”の正論理ボリュームAを取得し、ペア状態を“コピー中”に変更
ステップ 1010
正論理ボリュームAが存在するか？
Ｎｏ
処理終了
ステップ 1030
Ｙｅｓ
ステップ 1020
ジャーナル作成処理
ステップ 1040
コピー済みアドレスを変更
ステップ 1050
コピー済みアドレスが、正論理ボリュームの容量と等しいか？
Ｎｏ
Ｙｅｓ
ステップ 1060
正論理ボリュームAのペア状態を“正常”に変更
ステップ 1070

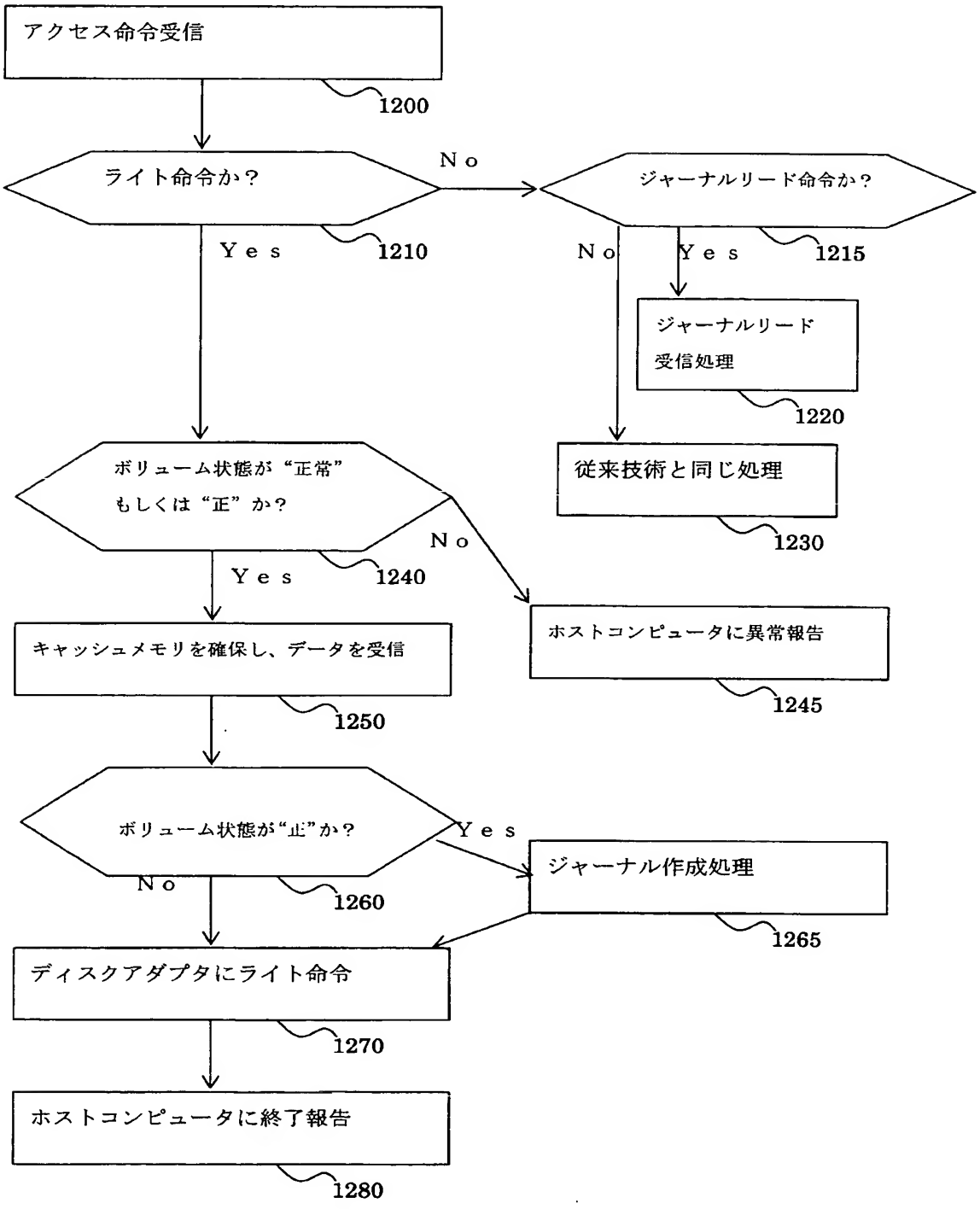
【図１２】

図１２


ホストコンピュータ
副記憶システム
180
100B
1100
命令受信処理
210
命令受信処理
210
1150
ホストアダプタ
110
1120
ホストアダプタ
110
更新情報
ライトデータ
キャッシュメモリ
130
1160
リードライト処理
220
リードライト処理
220
ディスクアダプタ
120
ディスクアダプタ
120
1110
1130
1140
正論理ボリューム
ジャーナル論理
ボリューム
230
230
正記憶システム
100A

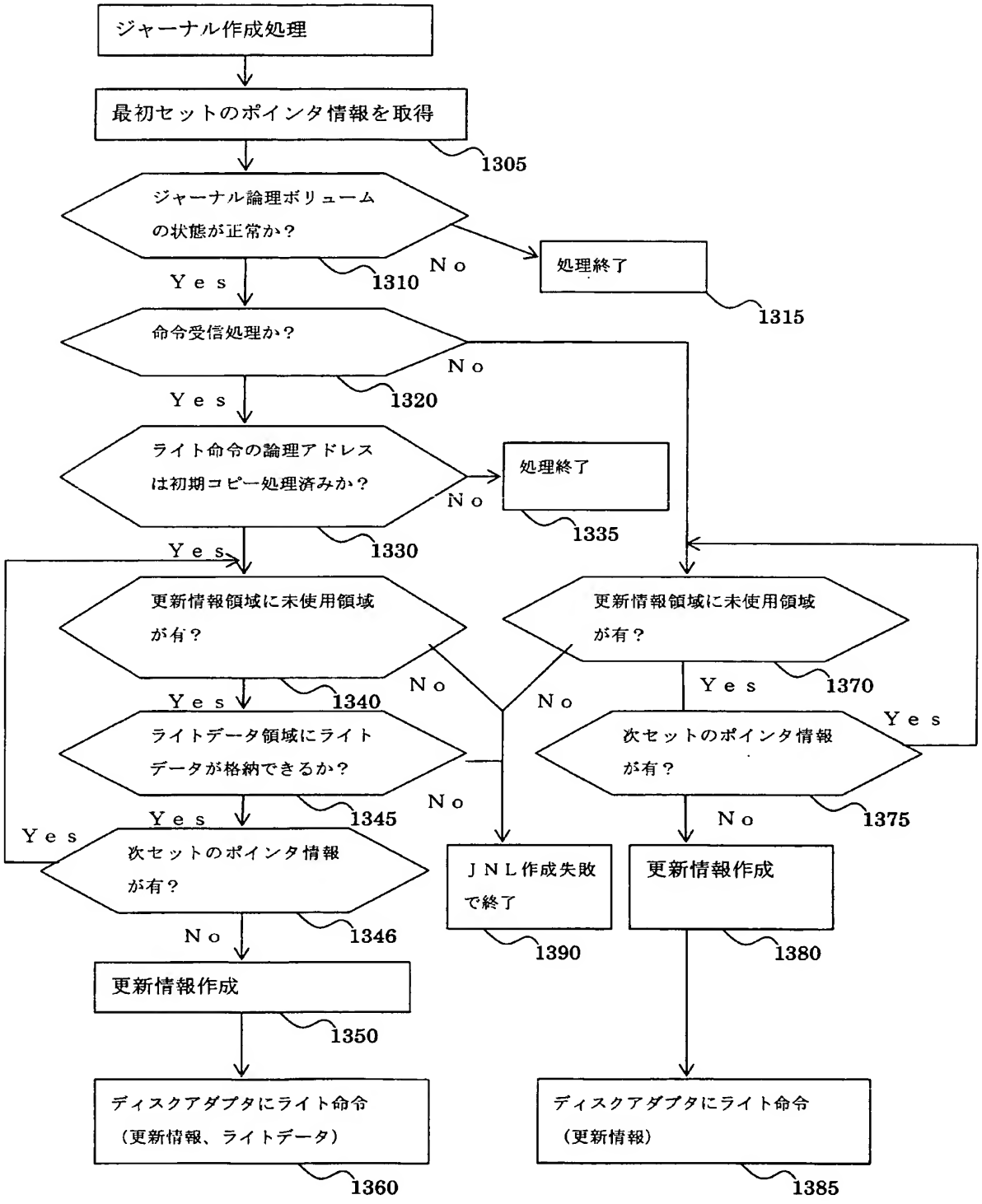
【図１３】

図１３


アクセス命令受信
1200
ライト命令か？
No
ジャーナルリード命令か？
Yes
1210
No
Yes
1215
ジャーナルリード
受信処理
1220
ボリューム状態が“正常”
もしくは“正”か？
従来技術と同じ処理
1230
No
Yes
1240
ホストコンピュータに異常報告
1245
キャッシュメモリを確保し、データを受信
1250
ボリューム状態が“正”か？
Yes
ジャーナル作成処理
No
1260
1265
ディスクアダプタにライト命令
1270
ホストコンピュータに終了報告
1280

【図１４】

図１４


ジャーナル作成処理
最初セットのポインタ情報を取得
1305
ジャーナル論理ボリューム
の状態が正常か？
No
1310
処理終了
1315
Yes
命令受信処理か？
No
1320
Yes
ライト命令の論理アドレス
は初期コピー処理済みか？
処理終了
No
1335
Yes
1330
更新情報領域に未使用領域
が有？
No
Yes
1340
ライトデータ領域にライト
データが格納できるか？
No
Yes
1345
次セットのポインタ情報
が有？
Yes
No
1346
更新情報作成
1350
ディスクアダプタにライト命令
（更新情報、ライトデータ）
1360
更新情報領域に未使用領域
が有？
No
Yes
1370
次セットのポインタ情報
が有？
Yes
No
1375
ＪＮＬ作成失敗
で終了
1390
更新情報作成
1380
ディスクアダプタにライト命令
（更新情報）
1385

【図１５】

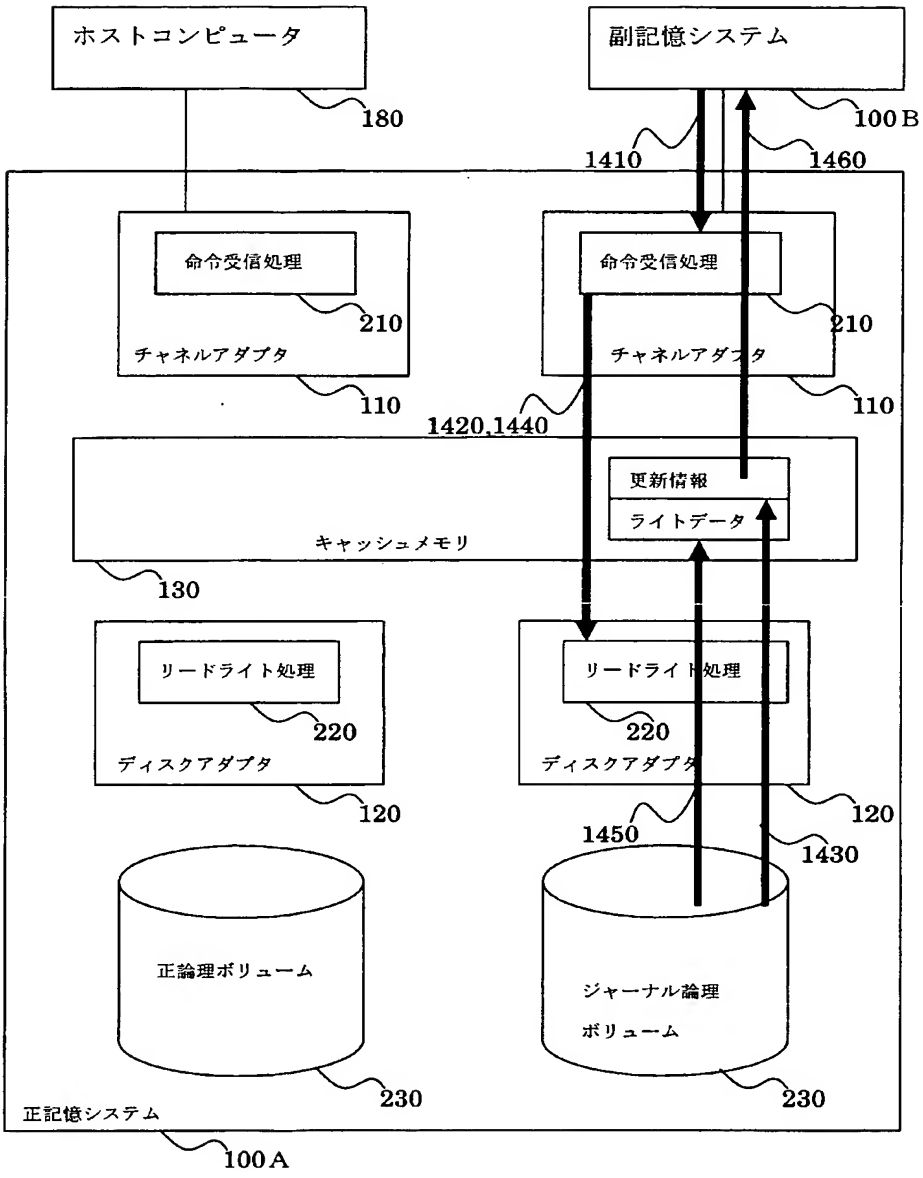
図１５

設定項目	設定値例
ライト命令を受信した時刻	1999/3/17 22:22:10
グループ番号	1
更新番号	4
ライト命令の論理アドレス	論理ボリューム番号：1 論理ボリュームの記憶領域の先頭からの位置：800
ライトデータのデータサイズ	100
ライトデータを格納したジャーナル論理ボリュームの論理アドレス	論理ボリューム番号：4 論理ボリュームの記憶領域の先頭からの位置：2200

310 更新情報

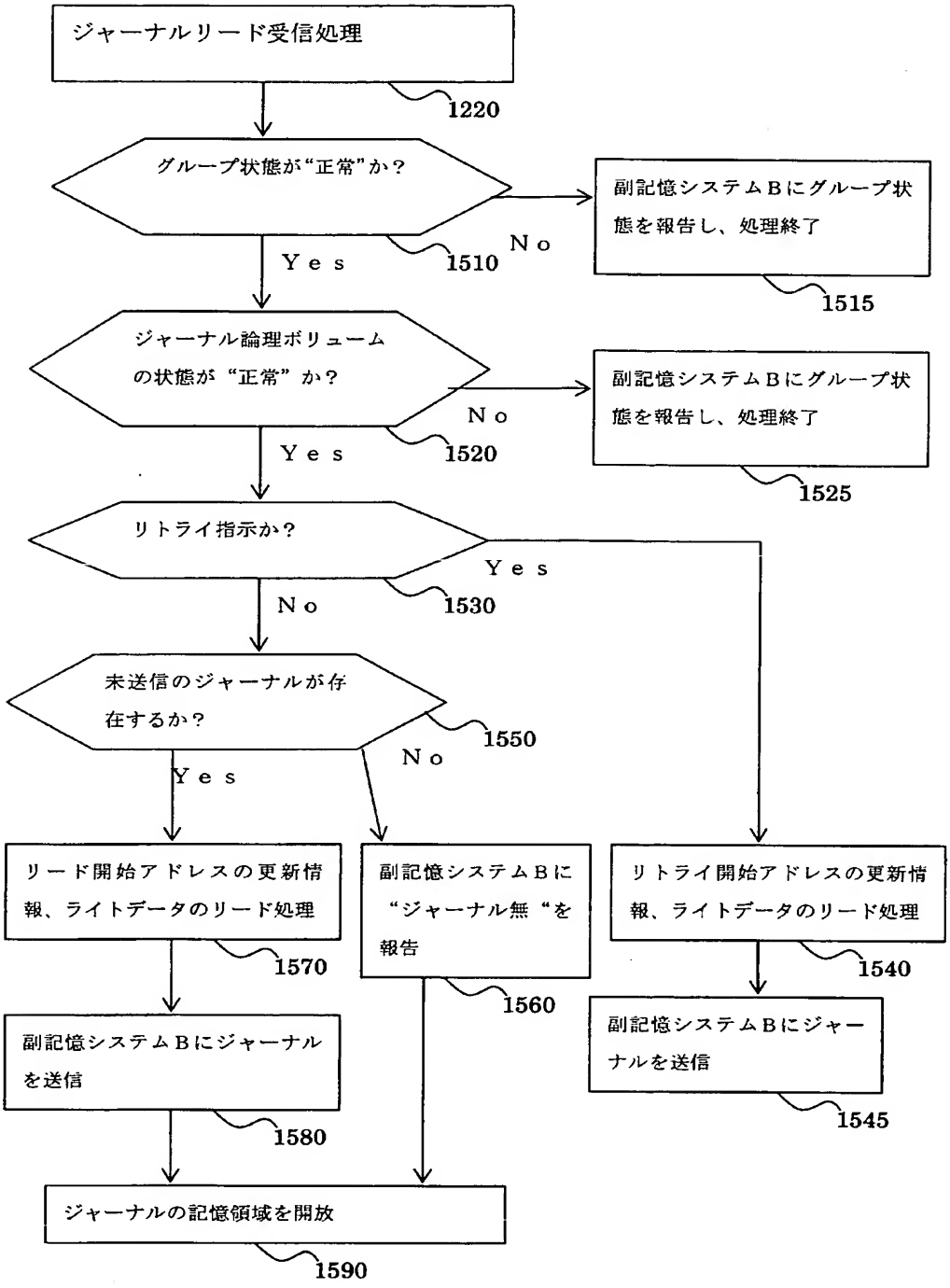
【図１６】

図１６


ホストコンピュータ
180
副記憶システム
100B
1410
1460
命令受信処理
210
チャネルアダプタ
110
命令受信処理
210
チャネルアダプタ
110
1420,1440
更新情報
ライトデータ
キャッシュメモリ
130
リードライト処理
220
ディスクアダプタ
120
リードライト処理
220
ディスクアダプタ
120
1450
1430
正論理ボリューム
230
ジャーナル論理
ボリューム
230
正記憶システム
100A

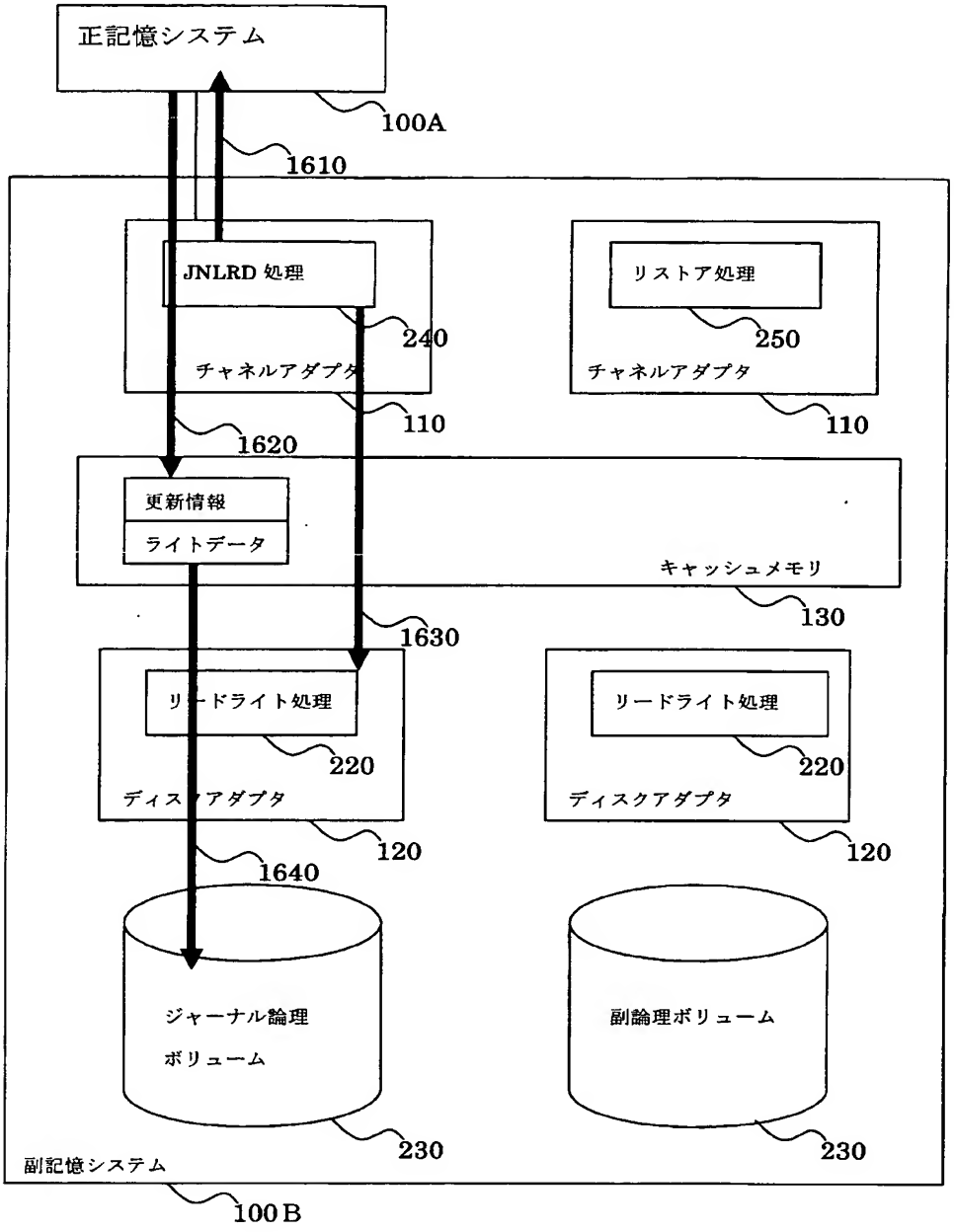
【図１７】

図１７


ジャーナルリード受信処理
1220
グループ状態が“正常”か？
1510
No
副記憶システムＢにグループ状態を報告し、処理終了
1515
Ｙｅｓ
ジャーナル論理ボリュームの状態が“正常”か？
1520
No
副記憶システムＢにグループ状態を報告し、処理終了
1525
Ｙｅｓ
リトライ指示か？
1530
Ｙｅｓ
No
未送信のジャーナルが存在するか？
1550
No
Ｙｅｓ
リード開始アドレスの更新情報、ライトデータのリード処理
1570
副記憶システムＢにジャーナルを送信
1580
副記憶システムＢに“ジャーナル無“を報告
1560
リトライ開始アドレスの更新情報、ライトデータのリード処理
1540
副記憶システムＢにジャーナルを送信
1545
ジャーナルの記憶領域を開放
1590

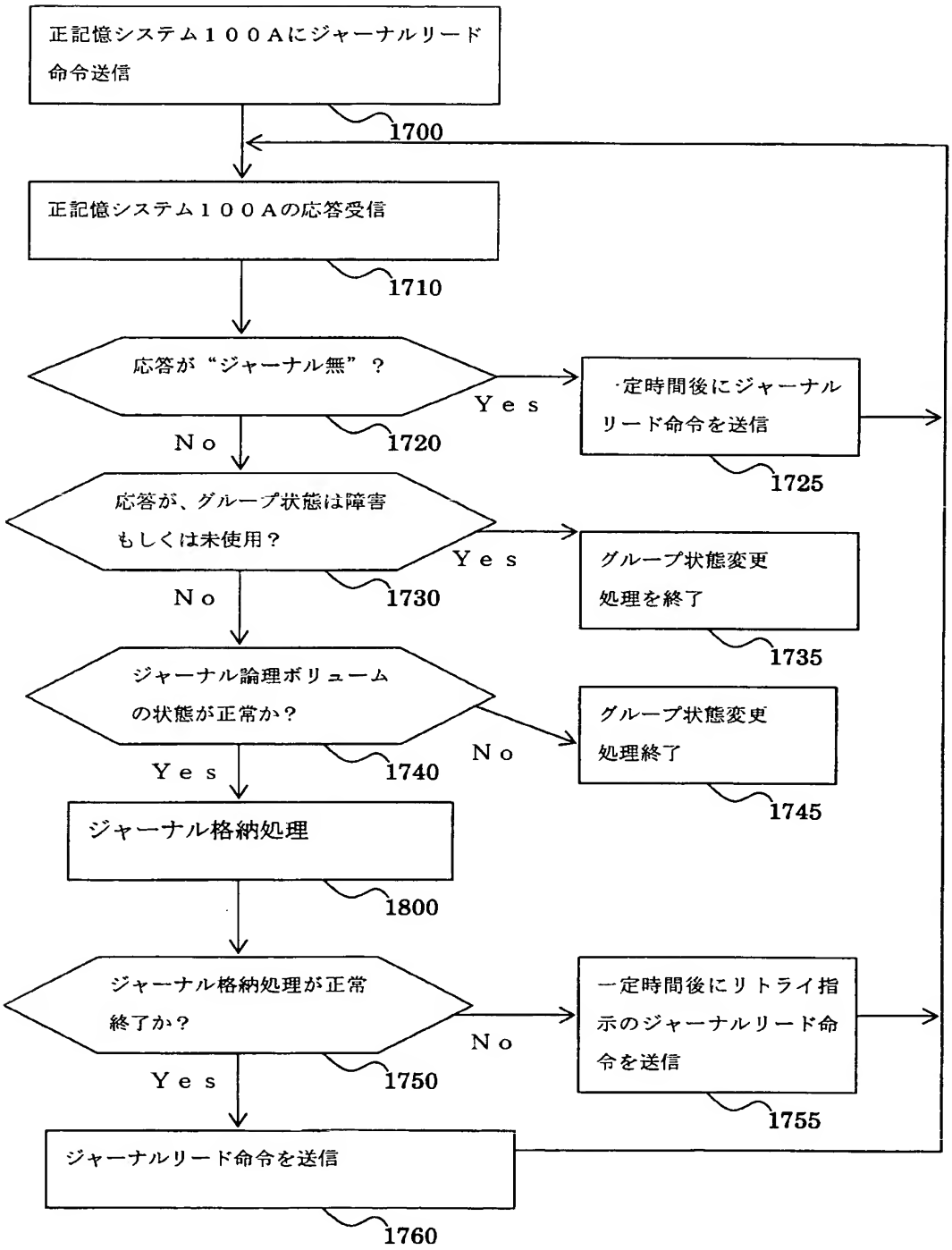
【図１８】

図１８


正記憶システム
100A
1610
JNLRD 処理
240
チャネルアダプタ
110
1620
リストア処理
250
チャネルアダプタ
110
更新情報
ライトデータ
キャッシュメモリ
130
1630
リードライト処理
220
ディスクアダプタ
120
1640
リードライト処理
220
ディスクアダプタ
120
ジャーナル論理
ボリューム
230
副論理ボリューム
230
副記憶システム
100B

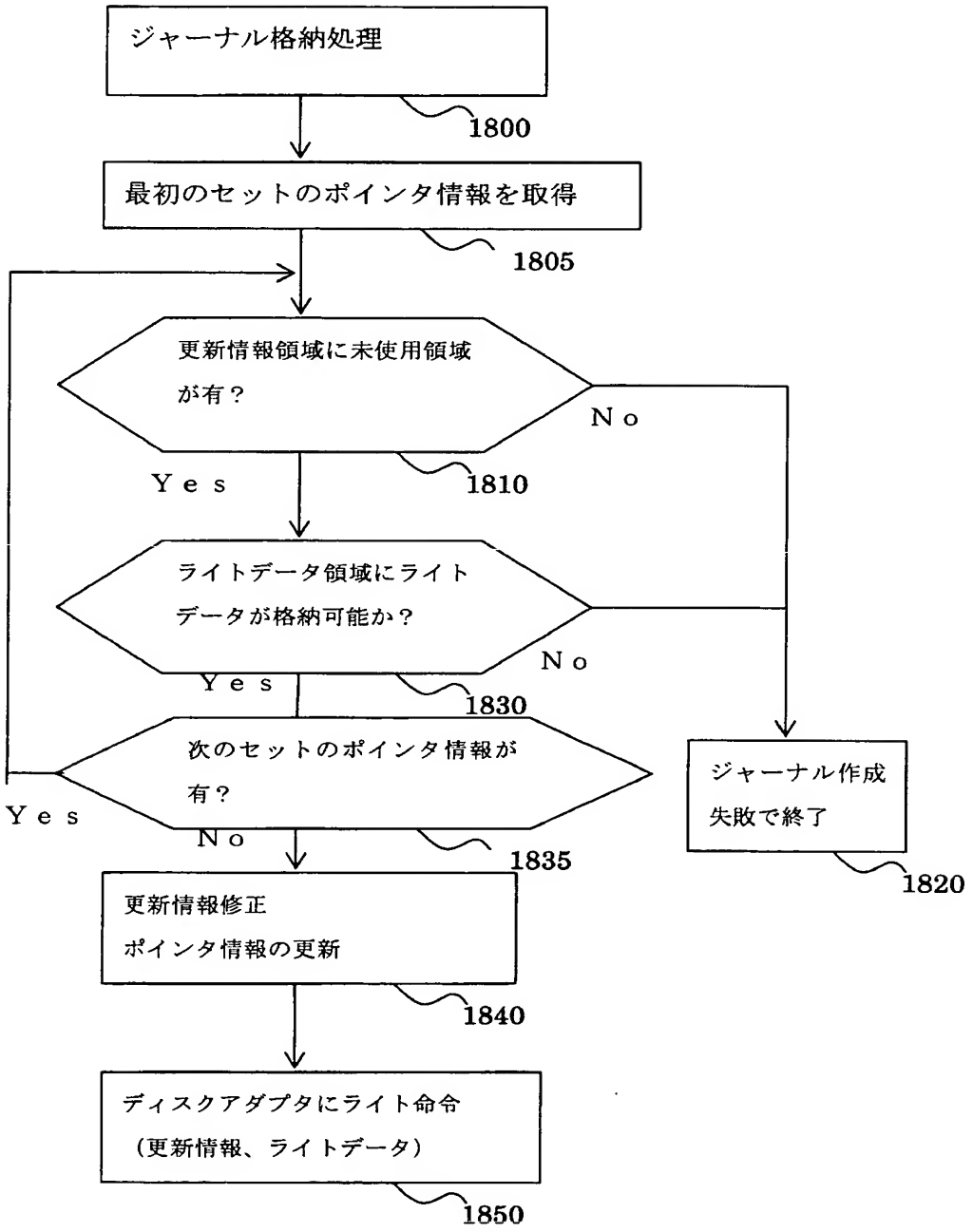
【図１９】

図１９


正記憶システム１００Ａにジャーナルリード命令送信
1700
正記憶システム１００Ａの応答受信
1710
応答が“ジャーナル無”？
Ｙｅｓ
1720
一定時間後にジャーナルリード命令を送信
1725
Ｎｏ
応答が、グループ状態は障害もしくは未使用？
Ｙｅｓ
1730
グループ状態変更処理を終了
1735
Ｎｏ
ジャーナル論理ボリュームの状態が正常か？
Ｎｏ
1740
グループ状態変更処理終了
1745
Ｙｅｓ
ジャーナル格納処理
1800
ジャーナル格納処理が正常終了か？
Ｎｏ
1750
一定時間後にリトライ指示のジャーナルリード命令を送信
1755
Ｙｅｓ
ジャーナルリード命令を送信
1760

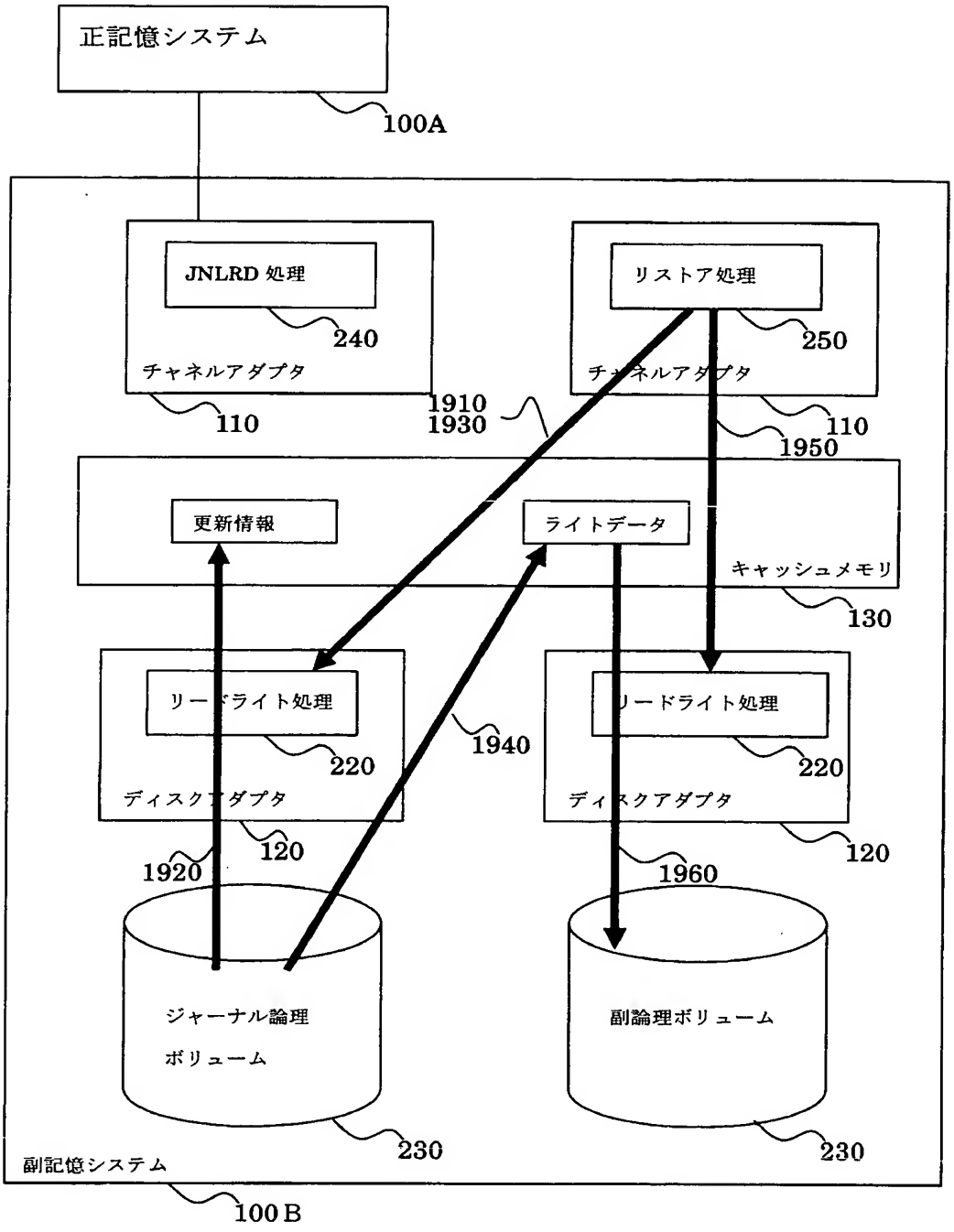
【図２０】

図２０


ジャーナル格納処理
1800
最初のセットのポインタ情報を取得
1805
更新情報領域に未使用領域が有？
No
Yes
1810
ライトデータ領域にライトデータが格納可能か？
No
Yes
1830
次のセットのポインタ情報が有？
Yes
No
1835
ジャーナル作成失敗で終了
1820
更新情報修正
ポインタ情報の更新
1840
ディスクアダプタにライト命令
（更新情報、ライトデータ）
1850

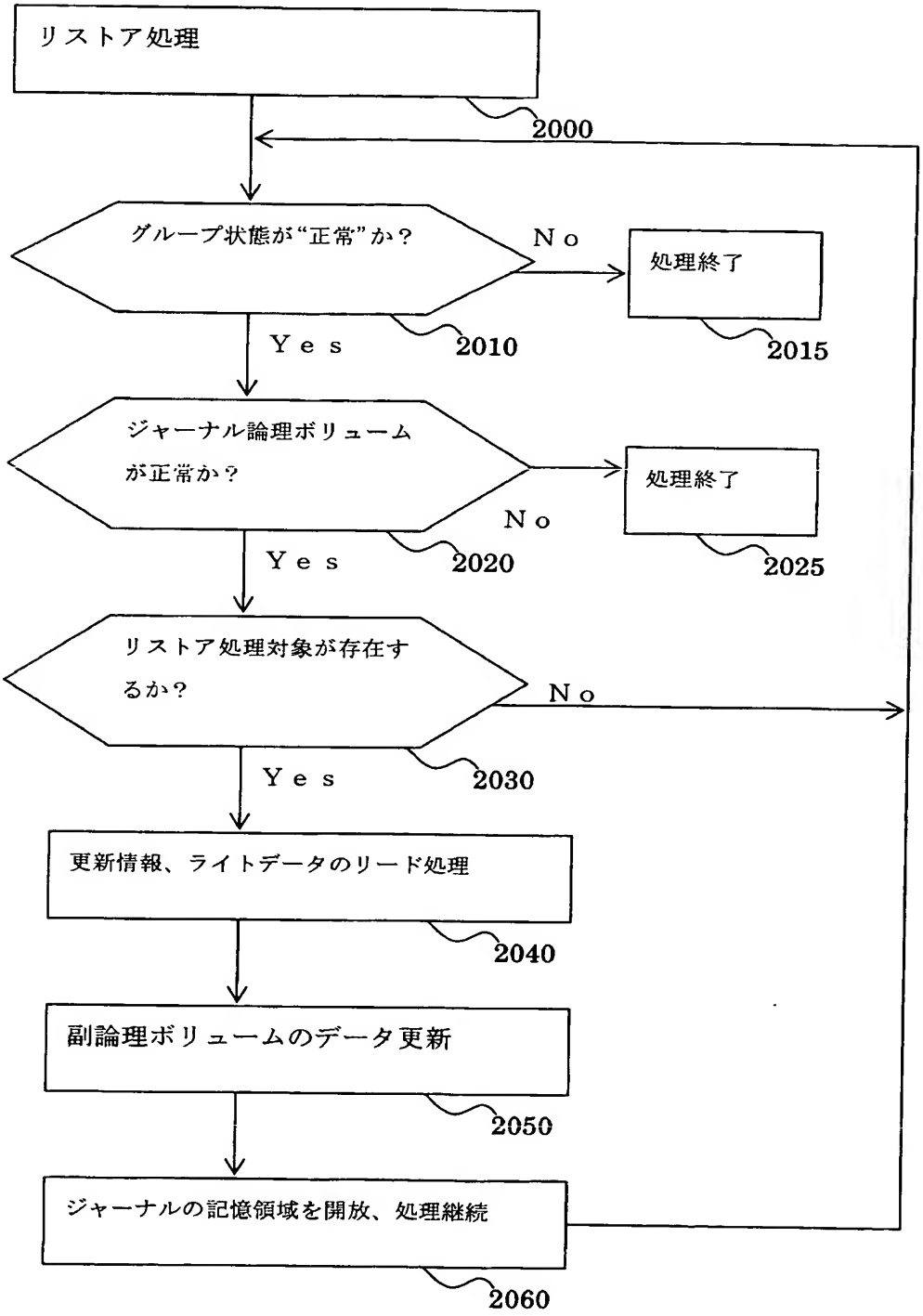
【図２１】

図２１

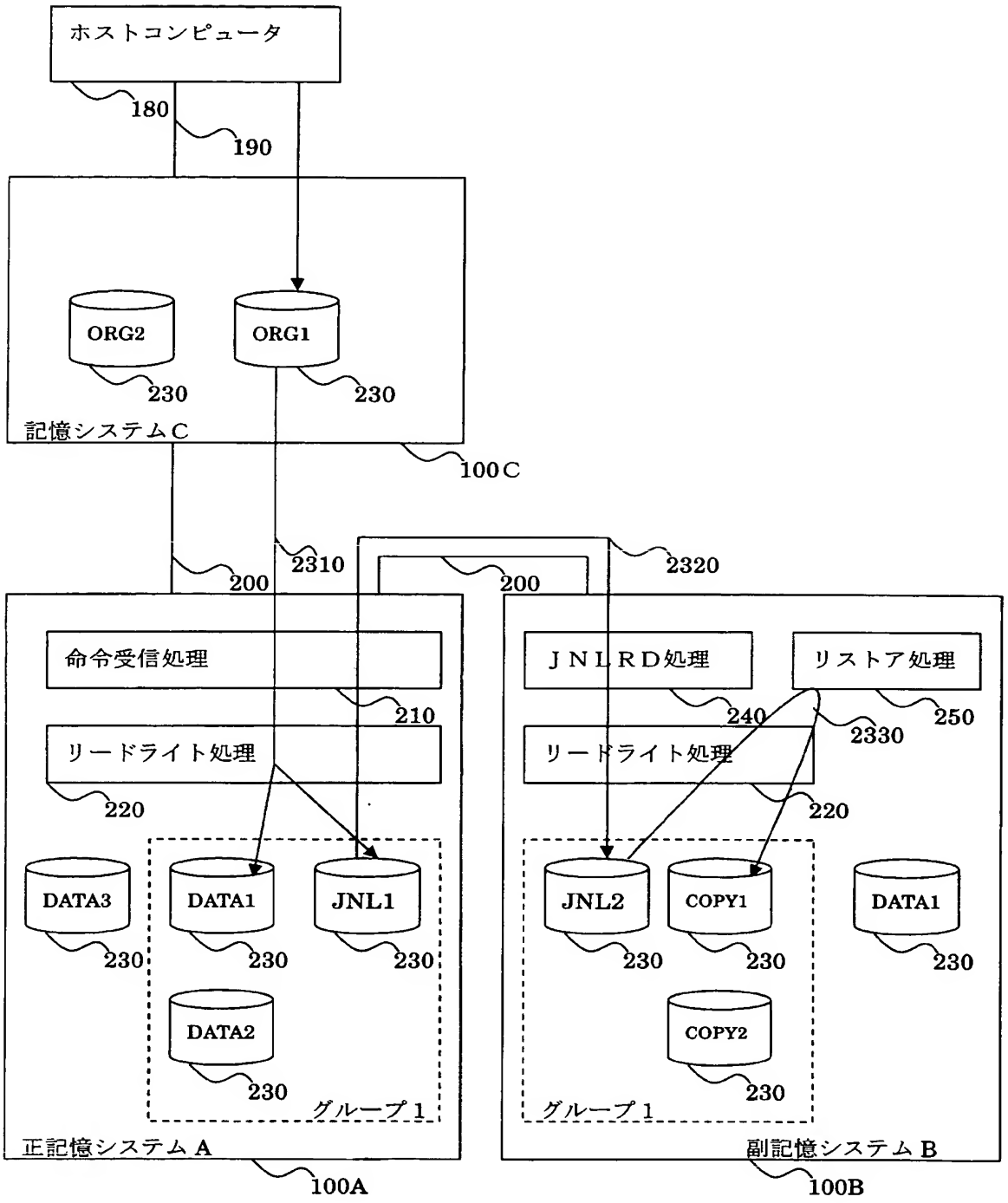

正記憶システム
100A
JNLRD 処理
240
チャネルアダプタ
110
リストア処理
250
チャネルアダプタ
110
1910
1930
1950
更新情報
ライトデータ
キャッシュメモリ
130
リードライト処理
220
ディスクアダプタ
120
1940
リードライト処理
220
ディスクアダプタ
120
1920
1960
ジャーナル論理
ボリューム
副論理ボリューム
230
230
副記憶システム
100B

【図２２】

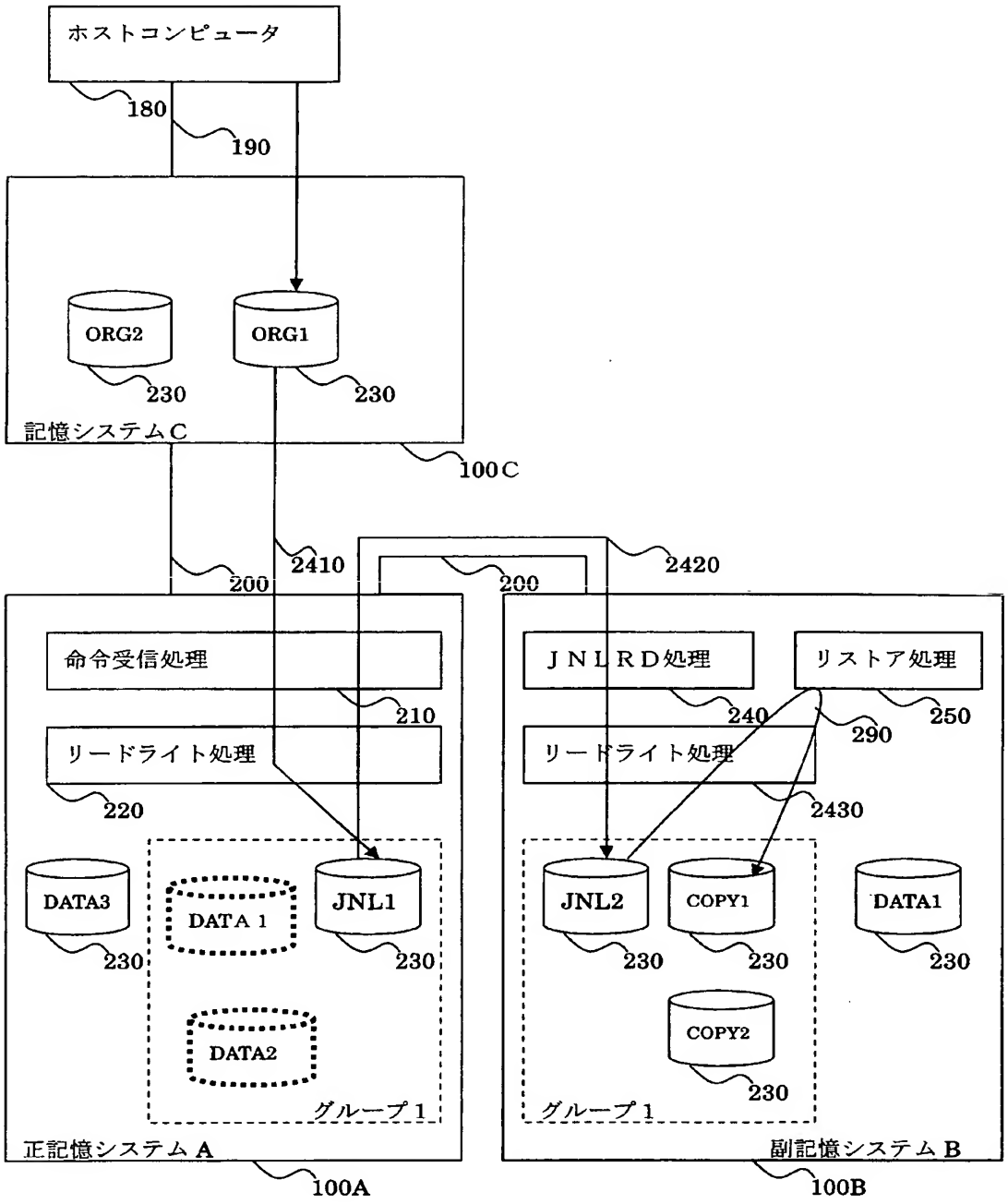
図２２


リストア処理
2000
グループ状態が“正常”か？
No
処理終了
2015
Yes
2010
ジャーナル論理ボリューム
が正常か？
処理終了
No
2025
Yes
2020
リストア処理対象が存在す
るか？
No
Yes
2030
更新情報、ライトデータのリード処理
2040
副論理ボリュームのデータ更新
2050
ジャーナルの記憶領域を開放、処理継続
2060

【図２３】

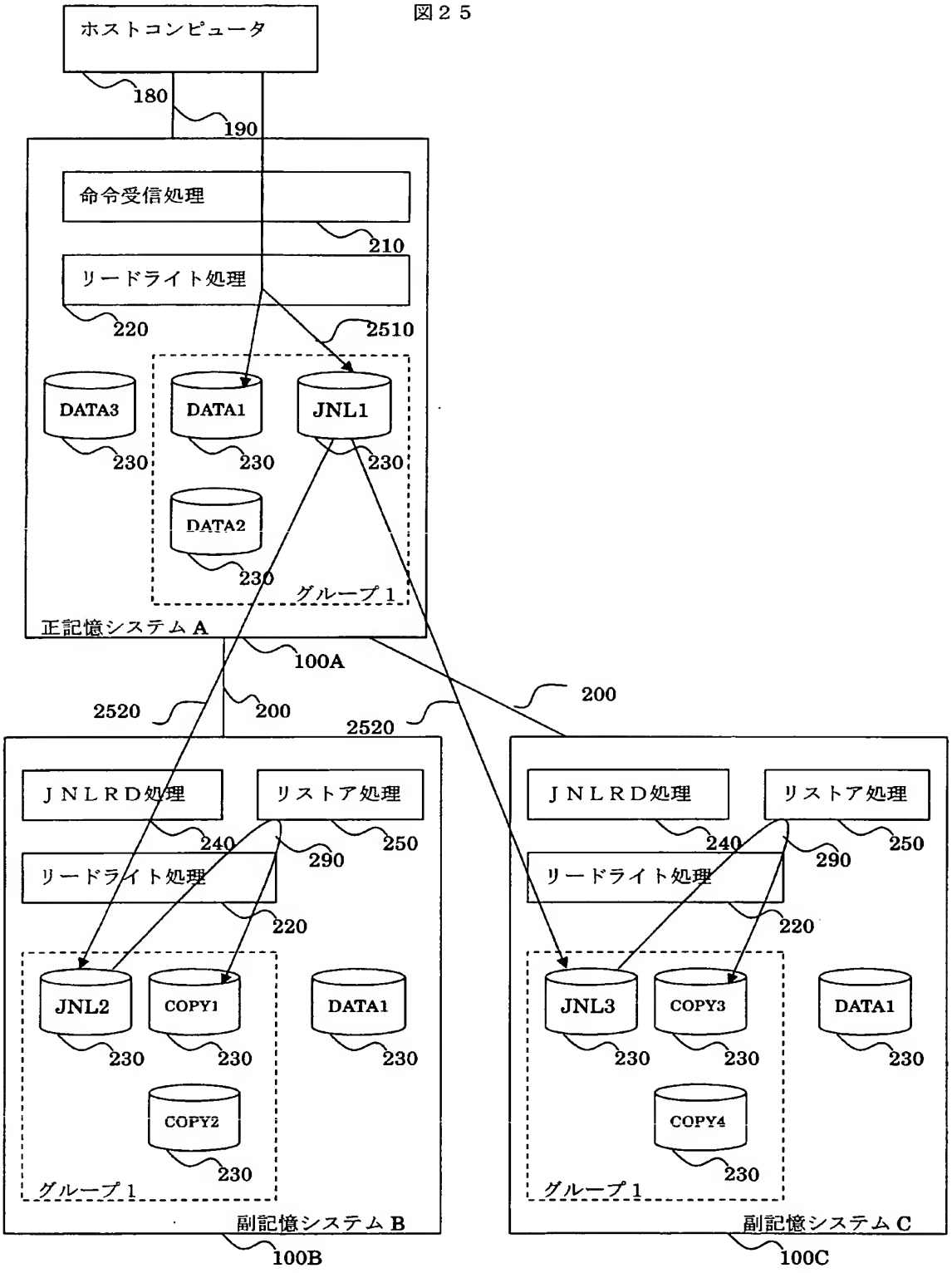

ホストコンピュータ
180
190
ORG2
ORG1
230
230
記憶システムC
100C
200
2310
200
2320
命令受信処理
210
リードライト処理
220
DATA3
DATA1
JNL1
230
230
230
DATA2
230
グループ1
正記憶システムA
100A
ＪＮＬＲＤ処理
リストア処理
240
250
2330
リードライト処理
220
JNL2
COPY1
DATA1
230
230
230
COPY2
230
グループ1
副記憶システムB
100B

【図２４】


ホストコンピュータ
180
190
ORG2
ORG1
230
230
記憶システムC
100C
200
2410
200
2420
命令受信処理
210
リードライト処理
220
DATA3
DATA 1
JNL1
230
230
DATA2
グループ1
正記憶システムA
100A
JNLRD処理
リストア処理
240
290
250
リードライト処理
2430
JNL2
COPY1
DATA1
230
230
230
COPY2
230
グループ1
副記憶システムB
100B

【図２５】

図２５


ホストコンピュータ
180
190
命令受信処理
210
リードライト処理
220
2510
DATA3
DATA1
JNL1
230
230
230
DATA2
230
グループ１
正記憶システムＡ
100A
2520
200
200
2520
ＪＮＬＲＤ処理
リストア処理
240
290
250
リードライト処理
220
JNL2
COPY1
DATA1
230
230
230
COPY2
230
グループ１
副記憶システムＢ
100B
ＪＮＬＲＤ処理
リストア処理
240
290
250
リードライト処理
220
JNL3
COPY3
DATA1
230
230
230
COPY4
230
グループ１
副記憶システムＣ
100C

【図２６】

図２６

	副記憶システム番号	論理アドレス	
		論理ボリューム番号	論理ボリュームの記憶領域の先頭からの位置
更新情報領域先頭アドレス	2	4	0
ライトデータ領域先頭アドレス	2	4	700
更新情報最新アドレス	2	4	500
更新情報最古アドレス	2	4	200
ライトデータ最新アドレス	2	4	2200
ライトデータ最古アドレス	2	4	1300
リード開始アドレス	2	4	400
リトライ開始アドレス	2	4	300

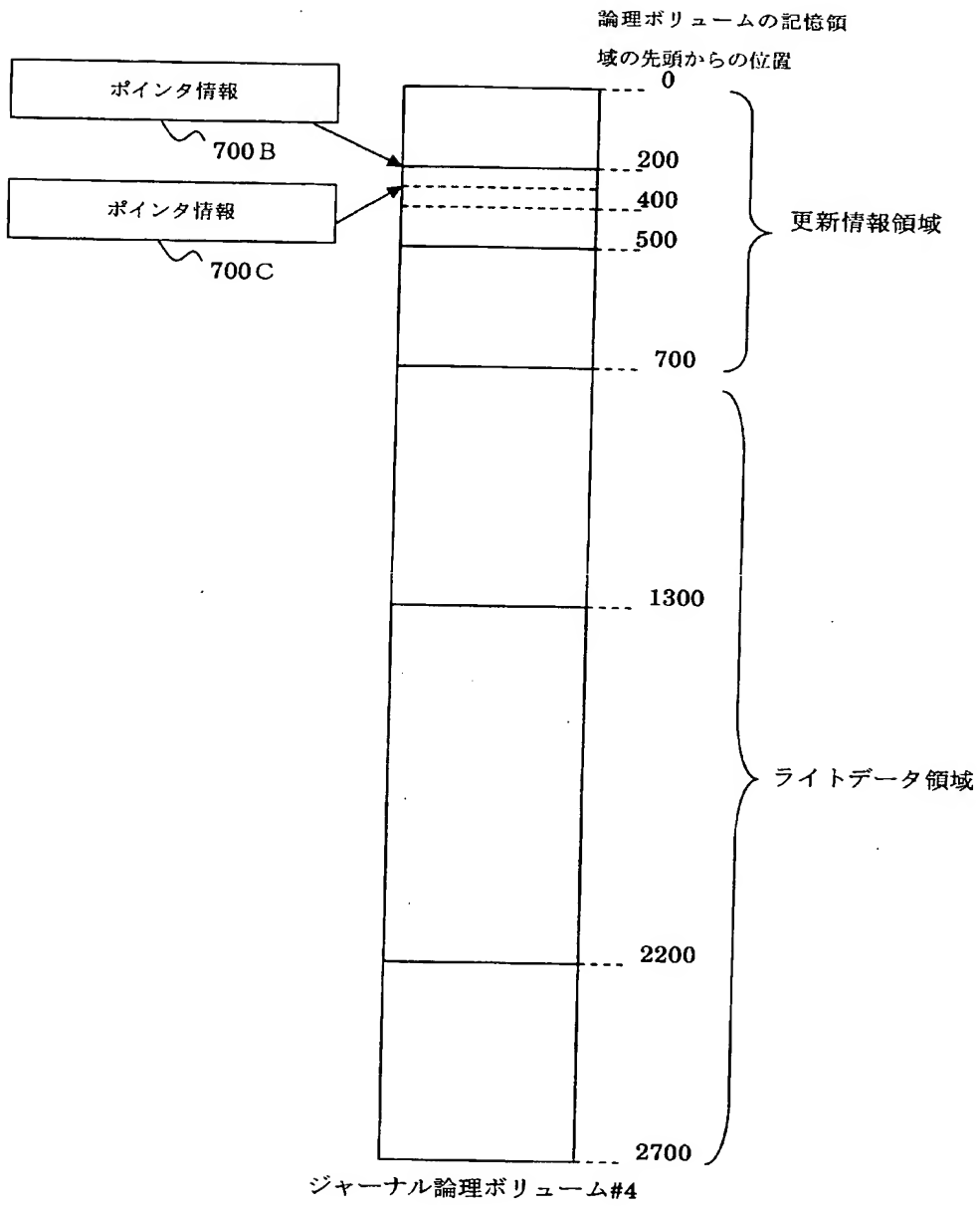
700B　ポインタ情報

	副記憶システム番号	論理アドレス	
		論理ボリューム番号	論理ボリュームの記憶領域の先頭からの位置
更新情報領域先頭アドレス	3	4	0
ライトデータ領域先頭アドレス	3	4	700
更新情報最新アドレス	3	4	500
更新情報最古アドレス	3	4	300
ライトデータ最新アドレス	3	4	2200
ライトデータ最古アドレス	3	4	1500
リード開始アドレス	3	4	500
リトライ開始アドレス	3	4	400

700C　ポインタ情報

【図２７】

図２７


論理ボリュームの記憶領域の先頭からの位置
ポインタ情報
700B
ポインタ情報
700C
0
200
400
500
700
1300
2200
2700
更新情報領域
ライトデータ領域
ジャーナル論理ボリューム#4

【書類名】要約書
【要約】
【課題】　記憶システムの上位の計算機に影響を与えず、複数の記憶システム間でデータ転送又はデータの複製を行う。
【解決手段】　上位装置１８０に接続された正記憶システム１００Ａに、２台以上の副記憶システム１００Ｂ、１００Ｃが接続される。副記憶システム１００Ｂ、１００Ｃは、それぞれ、独自のタイミングで正記憶システム１００Ａからデータ更新のジャーナルを読み所定の論理ボリュームＪＮＬ２、ＪＮＬ３に保存し、そして、独自のタイミングで論理ボリュームＪＮＬ２、ＪＮＬ３内のジャーナルに基づき、正記憶システム１００Ａ内のデータの複製を生成して、副論理ボリュームＣＯＰＹ１、ＣＯＰＹ３に保存する。正記憶システム１００Ａは、副記憶システム１００Ｂ、１００Ｃの双方がジャーナルを読んでリストアするまで、そのジャーナルを保持する。ジャーナルリードのタイミングはジャーナル量、処理負担などに応じて制御される。
【選択図】図２５

認定・付加情報

特許出願の番号	特願２００３－４１６４１４
受付番号	５０３０２０６０６５０
書類名	特許願
担当官	第七担当上席　　００９６
作成日	平成１５年１２月１６日

＜認定情報・付加情報＞

【提出日】　平成15年12月15日

特願２００３－４１６４１４

出　願　人　履　歴　情　報

識別番号　［０００００５１０８］

１．変更年月日　１９９０年　８月３１日
［変更理由］　新規登録
住　所　東京都千代田区神田駿河台４丁目６番地
氏　名　株式会社日立製作所